CHAPTER 2
REVIEW OF THE RELATED LITERATURE

The context of the research is first addressed by a review of various approaches for measuring the environmental impacts of human activities in Section 2.1. A premise of some of these approaches is that reducing environmental impacts and increasing economic efficiency can both be achieved. Production theory based methods are proposed as an alternative approach for measuring environmental performance at the manufacturing operations level of detail. Standard methods for the measurement of productive performance are reviewed in Section 2.2. The methods reviewed in Section 2.2 do not account for environmental impacts or environmental performance. In Section 2.3 various approaches for the measurement of performance are reviewed. In Section 2.4 the approaches already taken for the inclusion of environmental considerations in the measurement of productive performance are reviewed. Finally, in Section 2.5 the relevance of this literature is discussed in terms of the research.

2.1   Context of the Research

This research is focused on the application of methods for measuring productive efficiency of a single manufacturing facility. There are a number of related areas that provide a context for the research. The next five sections provide an overview of sustainable development, industrial ecology, pollution prevention and recycling, environmental economics, and life cycle cost analysis.

2.1.1 *Sustainable Development*

One commonly used definition of sustainable development is "to meet the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations, 1987). This general goal of sustainable development is as yet too ill-defined to be effectively incorporated into decisions at the global, national, local, and personal level. One key aspect of implementing sustainable development is the development

of metrics. Such metrics are needed "to provide a basis for social, political, economic and technical discussions relating to sustainability - tools to integrate environmental and sociopolitical conditions into market economics" (Pendergast, 1993).

The technological issue of efficient use of resources is only one part of what is required to achieve sustainable development. The so called IPAT equation developed by Paul Ehrlich and John P. Holdren captures the basic elements of human induced detrimental environmental impacts. "The impacts term (I) is a function of population (P), the level of affluence (A) and the technology (T) available" (Kates, 1994). This is expressed as:

$$I = f(P, A, T) \qquad\qquad (Eq. 2\text{-}1)$$

The focus of this research in on the third component, technology. Technology can be applied to reduce the impact of population and to allow a given level of affluence to be achieved with less environmental impact. Whether or not technological development alone can be assumed to overcome increasing population and increasing affluence to achieve sustainable development is a matter of controversy.

The assumption of technological progress as a means of achieving a sustainable civilization is dismissed as absurd by some, citing the fact that there is no evidence that this is the case other than "the past must also be a valid picture of the future" (Catton, 1980). Others are confident that technological progress and human creativity can readily support a projected human population of some 10 billion by the year 2050 (Bongaarts, 1994).

It is assumed that technological progress that increases the efficiency of resource use and the efficiency of the conversion of resource inputs to outputs is desirable. Whether or not such progress can allow sustainable development to occur without substantial changes in our civilization is outside the scope of this research.

With the increasing human impact on the environment, the resources devoted to environmental protection are increasing. Total expenditures for pollution abatement and

control in the United States are estimated to be \$87.564 billion for 1992 (U.S. Department of Commerce, 1994). Clearly, the protection of the environment has become a significant part of the U.S. economy. The efficient use of resources applied to pollution abatement and control is desirable both from the perspective of protecting the environment and maintaining competitive industries. The next three sections discuss approaches to achieving environmental goals in an efficient way.

2.1.2 *Industrial Ecology*

Industrial Ecology is a useful context for the development of approaches for achieving sustainable development. Industrial Ecology analogizes industrial systems to ecological systems. The complex-web of connections in ecological systems has evolved over time to essentially make use of all available energy and useful material. The waste of one organism is the food of another. In an industrial ecosystem wastes are not just outputs to be prevented but "also a part of the industrial process product stream that is to be designed" (Frosch, 1992, p. 800).

In an ideal Industrial Ecosystem there is maximum interaction among the major sectors of the economy. The result is the limited use of resources and the limited emission of wastes. One potential approach to increasing the interconnections among waste generators is an Eco-Industrial Park. An example of this kind of cooperation that can result in reduced impact on the environment as well as cost savings is provided by industries around the town of Kalundborg, Denmark. Cooperating industries are an electric power generating plant, an oil refinery, a biotechnology production plant, a plasterboard factory, a sulfuric acid producer, cement producers, local agriculture and horticulture, and district heating in Kalundborg. Among the 'waste' streams that are now being utilized are process steam from the electric power plant to the oil refinery, surplus gas from the oil refinery to the plasterboard factory, sulfur from the refinery to the sulfuric acid plant and many others (Tibbs, 1992). The experience of Kalundborg is an example of a large scale system integration problem that is part of the implementation of industrial ecology. In the United

14

States four such Eco-Industrial parks are in the planning stage (Chemical & Engineering News, 1995).

Implementation of industrial ecology and the achievement of environmental sustainability requires technological developments in two broad areas. One is the actual methods for reducing waste and recycling waste that is generated. Changes in processes and products are required. The other area is the need for "a coherent operational framework for examining potential long-term advantages and disadvantages of alternate webs of industrial changes and identifying the short-term bottlenecks that may emerge. These studies will provide the kind of information required both for public debate and decision making and for private calculations about requirements and opportunities. These debates, decisions, and calculations are necessary for the development of markets and as input to the various other social institutions that have a stake in industrial ecology" (Duchin, 1992, p. 851). This research is a contribution to the development of such an operational framework.

2.1.3 *Pollution Prevention and Recycling*

At the level of a particular facility or manufacturing process the application of Industrial Ecology takes the form of pollution prevention. The Pollution Prevention Act of 1990 defines Source Reduction to be any practice which:

> (i)   reduces the amount of any hazardous substance, pollutant, or contaminant entering any waste stream or otherwise released into the environment (including fugitive emissions) prior to recycling, treatment, or disposal; and
> (ii) reduces the hazards to public health and the environment associated with the release of such substances, pollutants, or contaminants (U.S. Congress, 1990).

Pollution prevention includes only source reduction. The EPA environmental management hierarchy places pollution prevention first, followed by recycling, treatment, and finally disposal in order of preference (EPA, 1992, p. 5).

In some organizations a pollution prevention program is undertaken in the wider context of a Total Quality Environmental Management program. A Total Quality Environmental Management program attempts to incorporate prevention of environmental problems into all operations as a means of achieving competitive advantage (GEMI, 1992). The same kind of paradigmatic change that is required as part of a Total Quality Management (TQM) program is also required as part of a pollution prevention program or what may be termed an Environmental TQM program (Green, 1993). An example of a related industry initiative is The Responsible Care Program of the Chemical Manufacturers Association that is intended to improve health, safety, and environmental performance. One key element of this program is pollution prevention that stresses continuous reductions in the amount of waste generated (Ember, 1995).

Although not generally considered to be part of pollution prevention activities, recycling is another means of reducing environmental impact without resorting to expensive end-of-pipe pollution controls. Recycling reduces environmental impact by reducing waste discharged or disposed and also by reducing the need for raw materials. Within each of the components of the industrial ecosystem as well as at particular facilities recycling is desirable as a means of reducing environmental impact. Recycled materials can also often be processed with less environmental impact than the virgin materials that are replaced. It has been estimated that "air emissions are reduced by 20 to 95% when scrap feedstocks replace virgin ones in various manufacturing processes. The range for water pollution reductions is 35 to 97 percent" (Institute for Local Self-Reliance, 1994, p. 12).

Many companies are now recognizing that pollution prevention can be a cost effective alternative to environmental control. DuPont has applied the three Rs of Manufacturing: Recycle, Reuse, and Reduce Waste at its plants. At a DuPont nylon factory in Chattanooga Tenn. the three R's have been applied to increase yields that results in less leftover materials, diverting waste streams to recycling, and then appropriately treating remaining waste streams (Lee, 1996).

Freeman (1992, p. 651) reviews industrial pollution prevention and cites many examples of successful programs and notes that how one measures "pollution prevention is the biggest P2 [pollution prevention] issue currently facing the industrial community." At the level of pollution prevention the concern is to appropriately assign costs and savings. Rooney (1992) presents an approach for estimating the cost of waste and notes that "excess expense caused by waste [can] vary from one-half to four times direct labor expense" (Rooney, 1992, p. 250). Three specific examples are provided in Table 2-1.

Table 2-1. Cost of Waste at Three Plants

	Cost of Waste	Direct Plant Labor
Fiberglass Pipe Plant	$1.443 million/yr	$1.837 million/yr
Pressed Plastic Parts Plant	$4.296 million/yr	$2.746 million/yr
Painting Fascias Plant	$23.315 million/yr	$6.65 million/yr

The cost of waste is calculated based on labor cost, disposal cost (includes material handling), and material cost. Material costs in the three examples range from 86.8% to 96.7% of the Cost of Waste. Table 2-2 shows the cost component of the cost of waste and the form this cost takes in standard costing systems (Rooney, 1992, p. 252). Rooney further divides material losses into nine categories presented in Table 2-3.

Table 2-2. Cost Components of Losses

Component of Cost of Waste	Form in Standard Costing System
Materials Losses	Materials loss allowances in standards, explicit and implied.
	Materials usage variances.
	Special loss accounts.
	Inventory loss or gain
Labor Losses	Labor in scrapped, reworked or otherwise discarded product.
Disposal Charges	Payments to outside disposal contractors from General Ledger.
Waste Handling Expense	Normally must be estimated from the number of staff employed in waste handling and reporting, using conservative waste and salary rates.

Table 2-3. Material Losses Cost Components

Category	Examples of Form
Underdelivery	Short weight in containers, tankers or rail cars
	Higher than standards water or solvent content in solutions or emulsions
	Handling Losses
	Burst or cut bags
	Material left in bags or drums
	Damaged Product
Process Losses	Set up pieces
	Chips machined off parts
	Sanding dust
	Off cuts, edge trim, offal
	Product left in mixing vessels, mills, tanks
Cleaning Losses	Solvent, resin or detergents used to clean equipment after production
Quality Losses	Scrap
	Customer Returns
	Rework
	Seconds
	Obsolescence
	Samples
Evaporation	Unexplained solvent loss
Stack Losses	Dust in bag houses
Overweight	Excess weight in packets
	Excess volume in containers
	Individual pieces weigh more than standard
Overspecification	Excess of expensive component in product
	Candy bar with higher than standard chocolate to center ratio
	Paint with higher than formula pigment
	Excess vinyl thickness on wallcovering stock

If the entire life cycle of a product is considered, further areas of potential pollution prevention can be identified. Table 2-4 provides a list of strategies for reducing environmental impact from the Electronics Industry Environmental Roadmap (1994, p. 65).

Table 2-4. Life Cycle Approaches for Reducing Environmental Impacts

Product Life Extension	Extend useful life
	Increased durability
	Ensure adaptability
	Increase reliability
	Expand service options
	Simplify maintenance
	Facilitate repairability
	Enable remanufacture of products
	Accommodate reuse of product
Material Life Extension	Develop recycling infrastructure
	Examine recycling pathways
	Use recyclable materials
	Material Selection
	Use substitute materials
	Devise reformations
Reduced Material Intensiveness	Conserve resources
	Process Management
	Process Substitution
	Process energy efficiency
	Process materials efficiency
	Process control
	Improved process layout
	Inventory control and material handling
	Facilities planning
	Treatment and disposal
Efficient Distribution	Optimize transportation systems
	Reduce packaging
	Use alternative packaging materials
Improved Management Practices	Using office materials and equipment efficiently
	Phase out high impact products
	Choose environmentally responsible suppliers
	Encourage labeling and advertise environmental claims

2.1.4 *Environmental Economics*

Environmental economics seeks to more appropriately account for the effects of pollution on the environment and on human beings. While the goal of environmental economics is consistent with the proposed research the approach differs. Environmental economics attempts to modify conventional economic theories and techniques to assign monetary value to non-market goods such as clean water and air so that market forces may produce the most efficient level of pollution control. This section reviews environmental economics and discusses why other approaches are useful.

There are two basic areas of research in environmental economics. One is focused on policy instruments for effecting reductions in pollution. A basic premise for much of this research is that polluters must pay a price equal to the marginal (i.e., additional) external cost of their polluting activities to induce them to internalize the full social cost of their activities (Cropper, 1992, p. 680). Research on policy options to taxing polluters include unit subsidies and marketable emission permits. Unit subsidies are based on the premise that a subsidy for emissions reduction can establish the same incentive for reducing pollution as a tax. In a permit trading option a permitting authority would issue permits that in the aggregate are equal to the efficient quantity of pollution and allow firms to bid for them (Cropper, 1992, p. 682).

The other main area of environmental economic research is in developing the costs and benefits of environmental amenities. This information is needed to implement some of the policy approaches for internalizing externalities to the polluter. Such information also may, in theory, allow tradeoffs between protecting the environment and other priorities such as economic development to be made.

One of the primary areas of research has been the development of methodologies to measure the benefits of goods such as clean water or air that are not sold on the market. The two basic approaches for making these estimates are indirect and direct market methods.

Indirect market methods attempt to infer tradeoffs between environmental goods and other goods from actual choices that people make. For example, inferences can be made on the value of environmental goods based on how much people are willing to pay for a house or where they choose to live. Direct market methods use surveys to ask people what tradeoffs they are willing to make between environmental and other goods (Cropper, 1992, p 700). Direct methods must link damage from pollution to something that people value (other than a clean environment). The something could be expected fish catch or visibility.

The techniques available for placing dollar values on improvements in environmental quality are limited and imprecise (Cropper, 1992, p. 729). All of these methods are limited by the fact that ecological systems and their response to pollutants are not well understood. Economic theory has developed based on private goods (i.e., goods that can be traded in a market). The theory of demand for public goods is an extrapolation from that existing theory (Green, 1991, p. 126). Assumptions made in economics may not be appropriate for the analysis of public goods. Some of these assumptions are:

- Adjustment to change is without cost and equilibrium is rapidly reached. It is assumed that any change, such as the imposition of a pollution tax, will immediately result in pollution reductions that represent the long term equilibrium. However, in many instances changes to the environment are irreversible making an equilibrium impossible to achieve. The loss of a species is an example of an irreversible change to the environment.

- The consumer possesses perfect information. In fact, knowledge of the impacts of human activities on the environment is limited. The public perception of the value of an ecosystem, that may be assigned an economic value, may not correspond to the ecological value (Green , 1991, p. 134).

- Because of discounting, benefits or costs occurring in the future can essentially be assigned a present value of zero. Benefits for future generations are essentially not considered. Yet many of the benefits and costs associated with environmental protection are long term.

- The theory of private goods assumes that society is a collection of individuals acting in their own self-interest (Green 1991, p. 127). This is clearly not a complete description of the way that society operates.

As a result, although environmental economics provides useful results for specific applications, useful guidance for the development of pollution controls, and useful methods for implementing pollution controls it is not yet up to the task of valuing ecosystems.

One approach for better integrating ecological considerations into the evaluation of economic systems is the emerging trans-disciplinary field of Ecological Economics. Ecological Economics takes sustainability as a goal and takes the view that humans are one component of an overall system (Costanza, 1991). Ecological Economics and Industrial Ecology are similar approaches in terms of goals and philosophy. Ecological Economics takes a more economics based approach while Industrial Ecology takes a more engineering based approach to achieving the common goal of sustainable development.

2.1.5   *Life Cycle Cost Analysis and Input-Output Analysis*

There are a number of approaches for the assessment of the environmental performance of products, processes, and management systems. Life Cycle Cost Analysis (LCA) is one of the primary methods that is being developed to help integrate environmental considerations into business decisions.

Keoleian (1994) reviews two components of environmental analysis for the evaluation of life cycle design. These are inventory analysis and impact analysis. "An inventory analysis identified and quantified inputs and outputs. In life cycle design, this inventory tracks materials, energy, and waste through each product cycle" (Keoleian, 1994, p. 99). Capital and labor are generally not tracked. Whenever possible, impacts are then characterized so different designs can be compared. To fully understand an impact, the pathways, fate, and effects of residuals (waste product from a waste treatment process) must be tracked. Such a full LCA is extremely time consuming and expensive given the large amount of data and analysis that is required. An abridged approach to LCA is suggested by Graedel, Allenby, and Comrie (1995) where a scoring system replaces rigorous estimation of

22

the costs for the LCA matrix elements. It is claimed that this approach "will identify perhaps 80% of the useful Design for Environment (DFE) actions that could be taken in connection with corporate activities, and the amounts of time and money consumed will be small enough that the assessment had a good chance of being carried out and its recommendations implemented" (Graedel, Allenby, and Comrie, 1995).

Some researches challenge the value of even performing such analyses. Lave, et. al. claim that "[e]qually credible analyses can produce qualitatively different results, so the results of particular life-cycle analysis cannot be defended scientifically. Thus current life-cycle analysis is not a reliable scientific tool" (Lave, et. al., 1995, p. 420A).

Another method that attempts to detail how inputs and outputs flow through the economy is Input-Output analysis (Leontief, 1986). This approach is strictly focused on commodity flows and does not address environmental impacts. This is an interesting approach from the perspective of studying the level of production system interactions to assess an industrial ecology. However, this method requires large amounts of data not typically available to a manufacturing facility. So, while Input-Output analysis may be usefully applied as a research method it is not appropriate for performance measurement at a manufacturing facility.

2.2　　The Measurement of Productive Performance

Although the methods previously reviewed all have value in certain contexts none of them are particularly useful for a single manufacturing facility. The proposed method, based on production theory, will use data typically available to a manufacturer to develop metrics of environmental performance. Any system that may be represented with inputs, outputs, and processes transforming inputs to outputs can be viewed as a production system. The productivity of such a system can be expressed as Output/Input. Increases in output with inputs held constant or decreases in inputs with outputs held constant result in increases in

productivity. There are many approaches to measuring productive efficiency. This section reviews the standard approaches that have been applied. These approaches may be divided into Frontier and Non-Frontier approaches. The Frontier based methods are reviewed in Section 2.2.1. Then, the axiomatic, set based notation and approach to the measurement of productive performance is reviewed in Section 2.2.2. This is the basis for the non-frontier methods based on dominance that are reviewed in Section 2.2.3. The method called Free Disposal Hull (FDH), while a frontier method, is based on definitions of dominance and so is discussed in Section 2.2.3 along with the non-frontier methods.

2.2.1 *Frontier Methods*

The measurement of technical change and the measurement of efficiency are based on the use of production functions or frontiers. Figure 2-1 illustrates the use of a production frontier to evaluate the efficiency of a production process. The isoquant or production frontier that represents the most efficient production process for transforming the inputs into a fixed output is shown in Figure 2-1 for a process with a single output and two inputs. Any production process that has a combination of inputs that may be represented by a point to the right of the production frontier uses more inputs for the same output and is therefore not efficient. Any point to the left of the production frontier uses less inputs for the same output than the most efficient possibilities and is by definition not possible for the particular production technology being evaluated.

Frontier based methods for measuring productivity first require that the production frontier or reference technology be defined. The two basic approaches to defining production frontiers are parametric (or econometric) and non-parametric (or mathematical programming). The mathematical programming approach is called Data Envelopment Analysis (DEA). The principle advantage of the econometric approach over the mathematical programming approach is that the effects of noise can be distinguished from the effects of inefficiency (C. A. Knox Lovell, 1993, p. 19). The principle advantage of the mathematical programming approach is that a functional form for the production frontier

need not be assumed. In a comparison of the DEA approach and parametric approach using a translog function it could not be conclusively stated that one measure of efficiency was superior to another, but the DEA method did provide more flexibility in terms of the shape of the production frontier (Banker, 1986, p. 42). The next section outlines the parametric approach, followed by a discussion of the DEA approach.
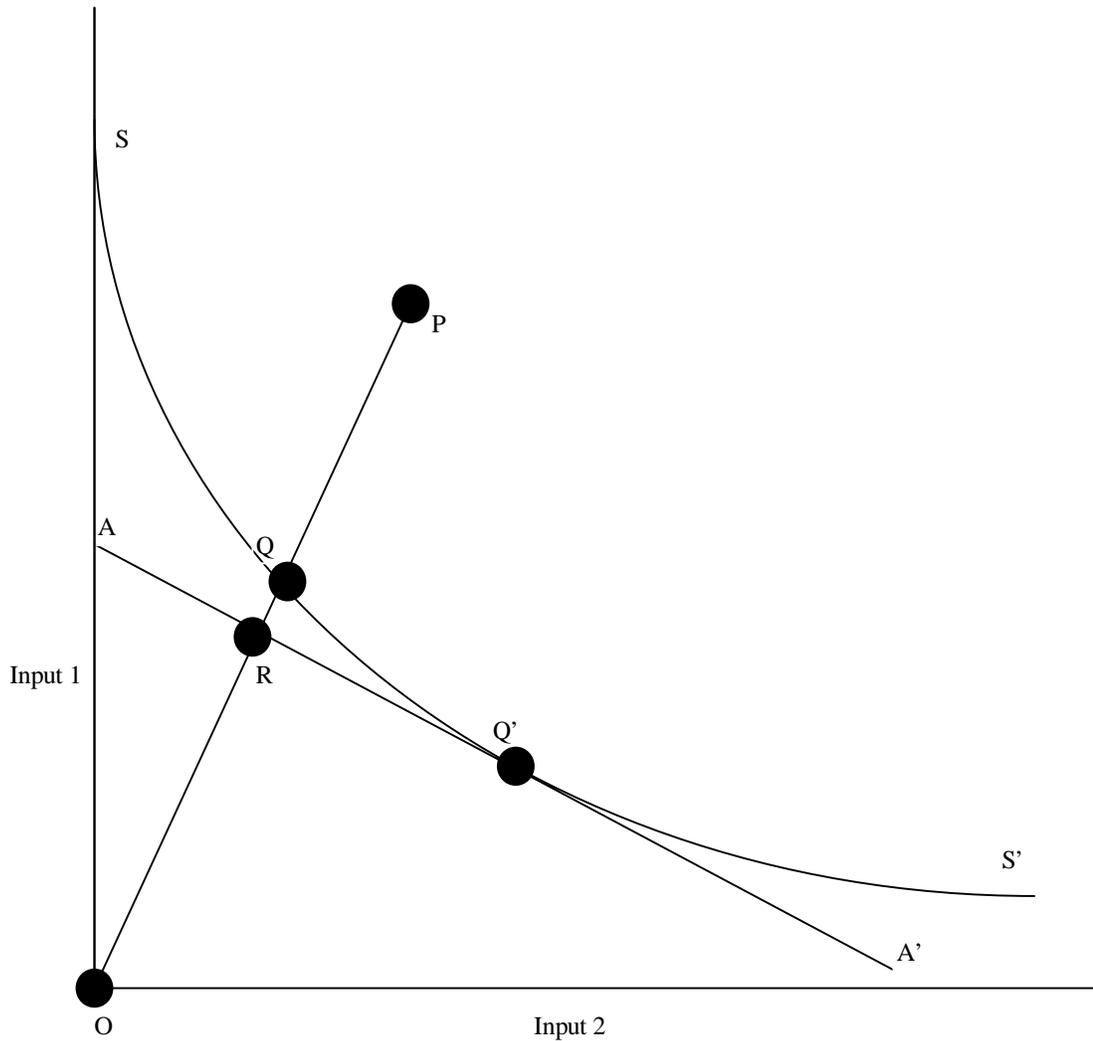

Figure 2-1. Farrell Technical and Economic Efficiency
(Farrell, 1957)

2.2.1.1 *DEA Approach*

Debreu (1951, p. 274) defines a distance function to be: "[T]he minimum distance from the given complex [production possibility] to a varying optimal complex [reference technology]." Debreu (1951, p. 285) defines the economic loss measured by the distance function to be composed of underemployment of resources, inefficiency in production, and imperfection of economic organization (i.e., monopoly). Using a distance measure, Farrell (1957) defines three measures of efficiency: technical, price, and overall (a combination of technical and price) that are based on a radial distance. These three basic measures of productive efficiency are shown in Figure 2-1.

For the same output, point P represents the inputs actually required and point Q represents the inputs required if production were as efficient as possible. The Technical efficiency of the production system represented for P is then defined to be OQ/OP. Technical efficiency is a measure of how well a production system minimizes inputs for a given level of output (as shown in Figure 2-1) or maximizes outputs for a given input. This may also be called resource efficiency.

There is also the question of what point on the production frontier is the most efficient combination of inputs for given prices. In Figure 2-1 the line AA' has a negative slope equal to the ratio of the prices of the two inputs. Point Q' where the line AA' is tangent to the production frontier is the optimal combination of inputs to produce the fixed output at the lowest cost. The fraction OR/OQ is defined as the price or allocative efficiency. Price efficiency is a measure of the success of a production system in choosing an optimal set of inputs for given prices. The combined price and technical efficiency, or overall efficiency, is defined by Farrell to be OR/OP. The Overall Efficiency is equal to the product of the technical efficiency and the price efficiency.

Technical efficiency as defined by Farrell (1957) is formulated by Charnes, Cooper and Rhodes (1978) (CCR) in terms of a linear programming approach with multiple inputs

and outputs. The extreme points (i.e., most efficient points) in a data set define the production frontier with the efficiency of particular points in the production set being defined relative to the set of efficient points. This is the Data Envelopment Analysis (DEA) approach to the measurement of efficiency. The basic definition of efficiency is the maximum ratio of weighted outputs to weighted inputs for the particular Decision Making Unit (i.e., production plan, data point) being evaluated subject to the condition that similar ratios of every Decision Making Unit (DMU) in the sample be less than or equal to one. These weights are determined as part of the solution of the linear programming formulation. This is expressed as:

$$\max \quad h_o = \frac{\sum_{j=1}^{J} \theta_j u_{j0}}{\sum_{i=1}^{I} v_i x_{i0}}$$

(Eq. 2- 2)

subject to:

$$\frac{\sum_{j=1}^{J} \theta_j u_{jk}}{\sum_{i=1}^{I} v_i x_{ik}} \leq 1 ; \quad k = 1,...,K$$

$$\theta_j, v_i \geq 0; \quad j=1, ...,J \; ; i = 1, ...,I. \text{ (CCR, 1978, p. 430)}$$

Where,

u_{jk} is the jth output for decision making unit k.

x_{ik} in the ith input used by decision making unit k.

$k = 1,…, K$ are Decision Making Units or production plans

$i = 1, ...I$ are the inputs.

$j = 1, ... J$ are the outputs.

θ_j, v_i are weights to be determined.

This non-linear formulation is then shown to be equivalent to an ordinary linear programming formulation making computation for large numbers of data points, k, tractable. (Bussafiane and Thanassoulis, 1991) This formulation is:

$$\max z_0 \qquad\qquad\qquad\qquad (Eq.\ 2\text{-}3)$$

subject to:

$$-\sum_{k=1}^{K} u_{jk}\lambda_k + u_{j0}z_0 \le 0; \qquad j = 1,...,J$$

$$\sum_{k=1}^{K} x_{ik}\lambda_k \le x_{i0}; \qquad i = 1,...,I$$

$$\lambda_k \ge 0; \qquad\qquad\qquad k = 1,...,K.$$

Since this is a ordinary linear programming formulation, the equivalent dual formulation is:

$$\min g_0 = \sum_{i=1}^{I} \omega_i x_{i0} \qquad\qquad\qquad (Eq.\ 2\text{-}4)$$

subject to:

$$-\sum_{j=1}^{J} \theta_j u_{jk} + \sum_{i=1}^{I} \omega_i x_{ik} \ge 0,$$

$$\sum_{j=1}^{J} \theta_j u_{j0} = 1,$$

$$\theta_j,\ \omega_i \ge 0$$

This formulation computes the output increasing measures of technical efficiency. The optimal solutions to these formulations of the linear programming problem are the same: $z_0 = g_0$ with $h_o = 1/z_0$.

2.2.1.2 Parametric Approach

The basic functional form for the parametric approach where the error term is decomposed is expressed as (C.A. Knox Lovell, 1993, p. 20):

$$u_k = f(x_k; \beta)e^{\{v_k + \theta_k\}}$$ (Eq. 2-5)

where:

u = scalar output

x = scalar inputs

k = 1,....,K production plans

v_k = term capturing statistical noise - assumed normally distributed about 0.

θ_k = nonpositive error component representing technical inefficiency.

β = vector of technology parameters to be estimated.

The econometric approach has many variations based on varying assumptions related to the distribution of the disturbance term, θ_k ; the basic model (i.e., whether the error term is added or multiplied to the function); the estimation method for the disturbance term (e.g., Ordinary Least Squares used in standard regression analysis) and parameter vector; and the measure of efficiency used (Kopp, 1981). The problem of assuming a specific form for a production frontier, with the resulting bias, still exists.

The distinction between parametric and non-parametric methods is now less pronounced due to recent work that seeks to overcome some of the disadvantages of the parametric approach to the measurement of efficiency. Banker (1992) retains the statistical error terms of the parametric approach, but does not explicitly assume a particular functional form. Instead, parametric assumptions are made about the inefficiency and random disturbance terms. The basic model used by Banker (1992) is:

$$u_j = u_j^f e^{-\theta_j} e^{vj}$$ (Eq. 2-6)

where,

u_j is an observed output

u_j^f is a frontier output

$e^{-\theta_j}$ is shortfall due to technical inefficiency

e^{vj} reflects random variations

The actual measure of efficiency is obtained in an approach similar to non-parametric methods in that linear programming is used to solve for a minimum difference between observed and the frontier output with constraints on the form of the frontier.

In yet another variation on the parametric approach, Thiry (1990) assumes a translog specification and then partitions the data set based on dominance. The dominant points are the ones that do not have any other point that has less input for some fixed output. In effect, the data set is partitioned between efficient points that are used to specify the parameters of the translog function and inefficient points. With the specified frontier, efficiency is measured based on the distance of a particular point to the specified production frontier.

With a model specified, there are a number of options for estimating the parameters of the production function. Standard regression, Ordinary Least Squares (OLS), methods may be used to estimate the production frontier. For a deterministic model[1], once the parameters are estimated by OLS, the production function may be shifted to so as to obtain a set of residuals all of which are the same sign (Green, 1993). This then represents the most efficient production function for the data. The production frontier may also be estimated from the Maximum Likelihood Estimates (MLE). For stochastic models MLEs are estimated based on various assumptions concerning the distributions of the efficiency error terms to include exponential, gamma, and half-normal (Green, 1993).

Recognizing that standard production and cost function techniques are limited by few available functional forms and restrictive properties a number of researchers have attempted to develop axiomatic, non-parametric approaches (Afriat, 1972) (Varian, 1984). Instead of assuming a functional form for the production function, the function is defined in terms of restrictions on its form such as concave, non-decreasing. Parameters for the production function are defined to maximize feasible output (or maximize or minimize other objectives such as profit and cost) by applying a mathematical programming approach to the

---

[1] A deterministic model assumes only random noise as a source of variation.

data. This approach is shown by Banker (1988) to be equivalent to the DEA approach of CCR in the sense that a set of data is partitioned into efficient points (upper bound) and inefficient points (lower bound).

The approach described in the next section assumes that all production occurs on the production frontier and therefore seeks only to measure technological progress or to compare the efficiency of different technologies.

2.2.1.3 *Technological Change*

Changes in technology capture technological progress while changes in efficiency capture the improved use of a particular technology. Technological progress results in the production frontier moving so that more output can be produced with the same inputs or that less inputs can be used to produce the same output. The measurement of technical change, however, typically makes the assumption that all production possibilities occur on the production frontier. Solow's (1957) treatment of the measurement of technological change is used to review the basic approach.

Proceeding on the assumption of neutral technical changes (i.e., marginal rates of technical substitution for inputs are constant) the production function takes the form (Solow, 1957):

$$U = A(t)f(X,L) \qquad\qquad\qquad\qquad \text{(Eq. 2-7)}$$

The factor $A(t)$ measures the cumulative effect of technological progress over time with $f(X,L)$ representing the production function for output U and inputs X and L. The calculation of $A(t)$ does not require that the parameters of the function $f(X,L)$ be known since the form of the production function is assumed constant and only movement of the production function is of interest. Solow uses a production function and its derivatives to show that the technological change component may be measured as a multiplier, $A(t)$. By differentiating with respect to time and dividing by U, Solow derives the following equation:

$$\frac{dU/dt}{U} = \frac{dA/dt}{A} + w_X \frac{dX/dt}{X} + w_L \frac{dL/dt}{L}$$

(Eq. 2-8)

where:

$$w_X = \frac{\partial X}{\partial K} \frac{X}{U} \text{ and}$$

$$w_L = \frac{\partial U}{\partial L} \frac{L}{U}$$

Richter (1966) shows that the ratio $\frac{dA/dt}{A}$ used by Solow as a gauge of technological progress is equivalent to the ratio $\frac{U(t)/U(0)}{N(t)/N(0)}$ where N is a Divisia Index. The Divisia index is a weighted sum of growth rates where weights are the component's share in total value (Hulten, 1973, p. 1017). Richter then modifies Solow's result to include Divisia Indexes for both inputs and outputs. This allows many outputs and inputs to be considered and results in symmetric treatment of the inputs and outputs.

The approach by Solow and the Divisia Index approach consider time to be continuous. The use of actual data, that is always discrete, requires approximation of the time derivative. Caves (1982a) develops a general framework for the use of discrete index numbers in the measurement of productivity. The Index Number approach to the measurement of technological change takes data (and time) to be discrete and is, therefore, better suited to the measurement of the technological change of actual production systems where data is only available at discrete time intervals. Caves attributes the basis for this approach to Malmquist:

> Malmquist in the consumer context, suggested comparing the input of a firm at two difference points in time in terms of the maximum factor by which the input in one period could be deflated such that the firm could still produce the output levels observed for the other time period. This results in a Malmquist input index, and there is an analogous Malmquist output index (Caves, 1982a, p. 1393).

32

Caves defines the Malmquist Index in terms of a distance function as:

$$U^0(x^1, x^0) = D^0(u^0, x^1)/D^0(u^0, x^0)$$
(Eq. 2-9)

Where 0 and 1 can either represent different times for a fixed technology or different technologies and where, x, are inputs and , u, are outputs.

Grosskopf (1992) relaxes the assumption of efficiency made by Caves to obtain an expression of the Malmquist productivity index decomposed into two components: one expressing technology change, one expressing inefficiency. Chambers, Färe, and Grosskopf (1994) further elucidates the relationship between measures of productivity using Index Numbers and Production Frontiers by defining the decomposed Malmquist Index in terms of distance functions as:

$$M_k(u^0, u^1, x^0, x^1) =$$
$$D_k^1(u^1, x^1)/D_k^0(u^0, x^0)\{[(D_k^0(u^1, x^1)/D_k^1(u^1, x^1)][D_k^0(u^0, x^0)/D_k^1(u^0, x^0)]\}^{1/2}$$

(Eq. 2-10)

Inputs, x, and outputs, u, are compared for two situations, (x^0, u^0) and (x^1, u^1). The ratio outside the bracket measures the Farrell technical efficiency. The bracketed term measures shifts in the frontiers. If it is assumed that all points must be on the frontier then the ratio $D_k^1(u^1, x^1)/D_k^0(u^0, x^0)$ is equal to one.

2.2.2 *Sets Representing Inputs, Outputs, and Production Technology*

The basic DEA approach previously described has been reformulated based on set theory. This research uses this set based approach and notation. Following Färe, Grosskopf, and Knox Lovell (1994) a production technology transforms inputs, **x**, to outputs, **u**. All possible inputs and outputs are part of the sets $\Re_+^I$ and $\Re_+^J$, respectively.

Inputs $\mathbf{x} = (x_1, x_2, ... x_I) \in \Re_+^I = \{\mathbf{x}: x \in \Re_+^I, x \geq 0\}$ (Eq. 2-11)

Outputs $\mathbf{u} = (u_1, u_2, ... u_J) \in \Re_+^J = \{\mathbf{u}: u \in \Re_+^J, u \geq 0\}$ (Eq. 2-12)

A production technology is represented by the output correspondence set P(x) or the input correspondence set L($\mathbf{u}$).

$$P(\mathbf{x}) = \{\mathbf{u} : \mathbf{x} \in L(\mathbf{u})\} \qquad \text{(Eq. 2-13)}$$

$$L(\mathbf{u}) = \{\mathbf{x} : \mathbf{u} \in P(\mathbf{x})\} \qquad \text{(Eq. 2-14)}$$

The Input Correspondence Set, L($\mathbf{u}$), represents all input vectors $\mathbf{x} \in \mathfrak{R}_+^I$ that yield at least output vector $\mathbf{u} \in \mathfrak{R}_+^J$. The Output Correspondence Set, P($\mathbf{x}$), represents all output vectors $\mathbf{u} \in \mathfrak{R}_+^J$ that are obtainable from the input vector $\mathbf{x} \in \mathfrak{R}_+^I$. Referring to Figure 2-1 the input set is represented by all input combinations to the right of the production frontier. For the case where inputs are fixed and outputs may vary, the output set is represented by all feasible outputs that may be obtained from the inputs.

The sets P($\mathbf{x}$) and L($\mathbf{u}$) represent the feasible sets of outputs and inputs that a particular technology may produce. The measurement of efficiency, however, depends on the definition of the efficient subsets of P($\mathbf{x}$) and L($\mathbf{u}$) that represents the production frontier or reference technology. For the case where outputs are constant and inputs may vary, shown in Figure 2-1, the production frontier is the subset of L($\mathbf{u}$) that represents the minimum inputs that may be consumed to produce the fixed output. Conversely, for the case where inputs are constant and outputs may vary, the production frontier is the subset of P($\mathbf{x}$) that represent the maximum outputs that may be produced given the fixed inputs.

Grosskopf (1986) reviews assumptions used to define reference technology in measuring productive efficiency. The following assumptions from Shephard (1981) are retained as a minimum requirement for the production frontier:

L.1 $\quad \mathbf{o} \notin L(u)\ (u > \mathbf{o})$ and $L(\mathbf{o}) = \mathfrak{R}_+^N,$ $\qquad$ (Eq. 2-15)

L.2 $\quad \mathbf{x} \in L(u) \Rightarrow \lambda \cdot \mathbf{x} \in L(u)\ (\lambda \geq 1),$ $\qquad$ (Eq. 2-16)

L.3 $\quad$ L is a closed correspondence, $\qquad$ (Eq. 2-17)

L.4 $\quad L(\Theta u) \subseteq L(u)\ (\Theta \geq 1)$ $\qquad$ (Eq. 2-18)

L.1 requires that some input is necessary for there to be an output and that factors of production can be idle. L.2 requires the weak disposability of inputs; meaning that output

cannot decrease when all inputs are increased proportionally. L.3 means that every point outside L(u) is in fact exterior to L(u) which guarantees that a minimum exits. L.4 is the assumption of monotonicity. A monotonic function is strictly increasing or decreasing. For the case of the input correspondence, L(u), this means an increase in output must occur with an increase in input or constant input. Constant, variable, and decreasing returns to scale are allowed by L.4.

Russell (1990) introduces the condition of commensurability (C) - i.e., the measure is independent of the units of measurements. This ensures that where measurement error exits that small errors of measurement result only in small errors of efficiency measurement. It is then show that (C) is inconsistent with a strict monotonicity requirement.

Once a reference technology is defined, the relative efficiency of a particular data point is measured by a distance function. Shephard (1981, p. 5-6) defines a production surface to be:

$$\psi (U, x_1, x_2,..., x_I) = 1 \tag{Eq. 2-19}$$

where,

ψ is a homogeneous function of degree one in $x_1, x_2,..., x_I$ for an arbitrary U.

U is one homogeneous output

$x_1, x_2,..., x_I$ are inputs for I factors of production

The intersection point of a ray passing through the origin on the production surface is denoted by $\xi (\xi_1, \xi_2,..., \xi_n)$. The distance function then is given by:

$$\psi = \left(\frac{\sum_{i=1}^{I} x_i^2}{\sum_{i=1}^{I} \xi_i^2} \right)^{\frac{1}{2}} \tag{Eq. 2-20}$$

The properties of the distance function are defined to be: (Shephard, 1981)

(a) $\psi (U,x) > 0$ for $x = (x_1, x_2,..., x_I)$ not coincident with the origin; (Eq. 2-21)

(b) $\psi (U, \lambda x) = \lambda \psi (U, \lambda x)$; (Eq. 2-22)

(c) if $x = (x_1,..., x_I)$ and $u = (u_1,...,u_J)$ are any to non-coincident points of the domain D (factor space), $\psi (U, x + u) > \psi (U, x) + \psi (U, u)$ (Eq. 2-23)

The actual measurement of efficiency requires that the production frontier be defined and that the distance that a particular production system is from the production frontier be defined. In the original presentation of DEA by CCR (1978) the relationship to Shephard's distance function is noted. Banker (1984) shows the equivalence between the CCR measure of efficiency and the reciprocal of Shephard's distance function with an assumption of free disposability, convexity and constant returns to scale.

2.2.3 *Dominance Based Methods*

The primary dominance based methods are Free Disposal Hull (FDH) and Benchmark Correspondence. The FDH method is a frontier method, but is discussed in this section since the definition of the frontier derives from the same definitions of dominance that are used to define the benchmark correspondence method.

2.2.3.1 *Benchmark Correspondence*

The Benchmark Production Correspondence method developed by Tulkens and Vanden Eeckaut (1991) does away entirely with a production frontier. The advantage of this is that progress as well as regress can be measured "without having to distinguish between progress and efficiency gains, or between regress and efficiency loss" (Tulkens and Vanden Eeckaut, 1991, p. 1). This method defines efficiency based on a pairwise dominance relation between production observations. Tulkens and Vanden Eeckaut (1995) specifically address the application of non-frontier measures of efficiency to time series. Since this is most applicable to the research, the Tulkens and Vanden Eeckaut (1995) paper is used in the rest of this section to describe the Benchmark Correspondence method, unless otherwise specified.

The basis of non-frontier methods (and Free Disposal Hull methods discussed in the next section) is the concept of dominance. Dominance, for pairwise comparisons, is defined by Tulkens and Vanden Eeckaut (1995, p. 85) as (Eq. 2-24):

Definition.   (1) $(\mathbf{x'},\mathbf{u'})$ dominates $(\mathbf{x},\mathbf{u})$ in inputs if both $\mathbf{u'} \geqq \mathbf{u}$ and $\mathbf{x'} \leq \mathbf{x}$.

(2) $(\mathbf{x'},\mathbf{u'})$ dominates $(\mathbf{x},\mathbf{u})$ in outputs if both $\mathbf{u'} \geq \mathbf{u}$ and $\mathbf{x'} \leqq \mathbf{x}$.

(3) $(\mathbf{x'},\mathbf{u'})$ dominates $(\mathbf{x},\mathbf{u})$ if either (1) or (2) holds or both.

The term, $\mathbf{x}$, is a I-dimensional vector of the inputs $(x_i, \ldots, x_I)$ and the term, $\mathbf{u}$, is the J-dimensional vector of outputs $(u_j, \ldots, u_J)$. The comparison is between the same elements of the input and output vectors of two production plans with one being designated as $(\mathbf{x'},\mathbf{u'})$ and the other as $(\mathbf{x},\mathbf{u})$. The term $\geqq$ means "greater than or equal to" and the term $\leqq$ means "less than or equal to." The term $\leq$ means "less than or equal to with at least one element of the vector $\mathbf{x}$ or $\mathbf{u}$ less than." The term $\geq$ means "greater than or equal to with at least one element of the vector $\mathbf{x}$ or $\mathbf{u}$ greater than." This definition of dominance is similar to one provided by Koopmans (1951, p. 460) where a production plan would be considered efficient "if there is no other attainable set of commodity flows in which all flows are at least as large as the corresponding flows in the original set, while at lease one is actually larger."

When evaluating time series data there are three basic approaches possible as described by Tulkens and Vanden Eeckaut (1995):

1. A single reference set based on all of the data - called intertemporal.

2. A reference set is defined for each time, t - called contemporaneous.

3. A reference set is defined for each time, t, based on all of the observations from the beginning of the observation period until time t - called sequential.

For time series data from a single production system, a contemporaneous reference set, is not meaningful. An intertemporal reference set may be constructed, but assumes a fixed reference set over time. Regress cannot be measured as distinct from inefficiency. The third

approach, sequential, allows both progress and regress to be measured. This is the approach taken by Tulkens and Vanden Eeckaut (1995) to define a Benchmark Correspondence for time series data.

The Benchmark Correspondence is a graph of some relation between the inputs and the output. The graph of this correspondence is the set $D_i(1)$ - the Benchmark Correspondence. A benchmark production correspondence at time t is defined to be:

$$B_t : \mathfrak{R}_+^I \to \mathfrak{R}_+^J ; x \mapsto U(x) = \left\{ u \big| (x,u) \in D_i(1,t) \right\} \qquad \text{(Eq. 2-25)}$$

The sets defined below are determined sequentially using time series data. Tulkens and Vanden Eeckaut (1991) define three sets of observations based on dominance at time t=1 for the whole space $\mathfrak{R}_+^{I+J}$ are:

$$D_d \ (1) \ = \ \left\{ \begin{bmatrix} u \\ x \end{bmatrix} \in \mathfrak{R}_+^{I+J} \left\| \begin{bmatrix} u \\ x \end{bmatrix} = \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} + \sum_{j=1}^J \mu_j \begin{bmatrix} O^I \\ e_j^J \end{bmatrix} - \sum_{i=1}^I v_i \begin{bmatrix} e_i^I \\ O^J \end{bmatrix}, \mu_j \geq O \forall_j, v_i \geq O \forall_i \right\} \setminus \left\{ \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} \right\}$$

(Eq. 2-26)

$$D_g \ (1) \ = \ \left\{ \begin{bmatrix} u \\ x \end{bmatrix} \in \mathfrak{R}_+^{I+J} \left\| \begin{bmatrix} u \\ x \end{bmatrix} = \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} - \sum_{j=1}^J \mu_j \begin{bmatrix} O^I \\ e_j^J \end{bmatrix} + \sum_{i=1}^I v_i \begin{bmatrix} e_i^I \\ O^J \end{bmatrix}, \mu_j \geq O \forall_j, v_i \geq O \forall_i \right\} \setminus \left\{ \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} \right\}$$

(Eq. 2-27)

$$D_i \ (1) \ = \ \left\{ \begin{bmatrix} u \\ x \end{bmatrix} \in \mathfrak{R}_+^{I+J} \left\| \begin{bmatrix} u \\ x \end{bmatrix} = \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} + \sum_{j=1}^J \mu_j \begin{bmatrix} O^I \\ e_j^J \end{bmatrix} + \sum_{i=1}^I v_i \begin{bmatrix} e_i^I \\ O^J \end{bmatrix}, \mu_j \geq O \forall_j, v_i \geq O \forall_i \right\} \setminus \left\{ \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} \right\}$$

$$\cup \left\{ \begin{bmatrix} u \\ x \end{bmatrix} \in \mathfrak{R}_+^{I+J} \left\| \begin{bmatrix} u \\ x \end{bmatrix} = \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} - \sum_{j=1}^J \mu_j \begin{bmatrix} O^I \\ e_j^J \end{bmatrix} - \sum_{i=1}^I v_i \begin{bmatrix} e_i^I \\ O^J \end{bmatrix}, \mu_j \geq O \forall_j, v_i \geq O \forall_i \right\} \setminus \left\{ \begin{bmatrix} u^1 \\ x^1 \end{bmatrix} \right\}$$

(Eq. 2-28)

where,

D_d is the set of points in the space that are weakly dominated by observation

(x^1, u^1)

D_g is the set of points in the space that are weakly dominating observation

$(\mathbf{x}^1, \mathbf{u}^1)$.

D_i is the set of points in the space that are dominance-indifferent with respect to observation $(\mathbf{x}^1, \mathbf{u}^1)$ i.e, neither dominating nor dominated by it.

e_i^J, e_i^I = J-dimensional and I dimensional zero vectors where the ith and the jth component, respectively, is equal to one.

O^I, O^J = I-dimensional and J-dimensional zero vectors.

μ_j, ν_i = scalars for all i and j.

i and j = Input and Output quantities, respectively.

The point $(\mathbf{x}^1, \mathbf{u}^1)$ is subtracted from the expression between braces as it cannot dominate itself.

Figures 2-2, 2-3, and 2-4 illustrate these sets. (Tulkens and Vanden Eeckaut, 1991, p. 11-12). The reference production plan, $(\mathbf{x}^1, \mathbf{u}^1)$, is shown at the center of each of the three figures. Other production plans are compared to this reference to determine what set, Dg, Di, or Dd in which to place the production plan being evaluated. The first production plan in a time series, $(\mathbf{x}^1, \mathbf{u}^1)$, is always placed in the set, Di, and is therefore also one of the reference production plans. Subsequent production plans are compared to this first production plan. Each production plan that is placed in the set, Di, then becomes a reference against which subsequent production plans are compared. This algorithm is described below.

(i) the first observation $(\mathbf{x}^1, \mathbf{u}^1)$ is always retained as a benchmark observation;

(ii) the second observation $(\mathbf{x}^2, \mathbf{u}^2)$ is retained as a benchmark observation if and only if it neither dominates nor is dominated by the previous observation (that was retained as a benchmark observation at time 1);

(iii) for the third and each one of the following observations $(\mathbf{x}^t, \mathbf{u}^t)$ t = 4, ...T, it is retained as a benchmark observation if and only if it neither dominates nor is dominated by any of the previous observations that were retained as benchmark observations at the time they were made.

These production plans make-up the benchmark correspondence that will yield the reference set for evaluating all observations. Since there is no frontier, efficiency is not measured. What is measured are indexes of progress and regress.

u

D_g

Dominating

D_i

Dominance Indifferent

u^1

D_i

Dominance Indifferent

D_d

Dominated

x^1

x
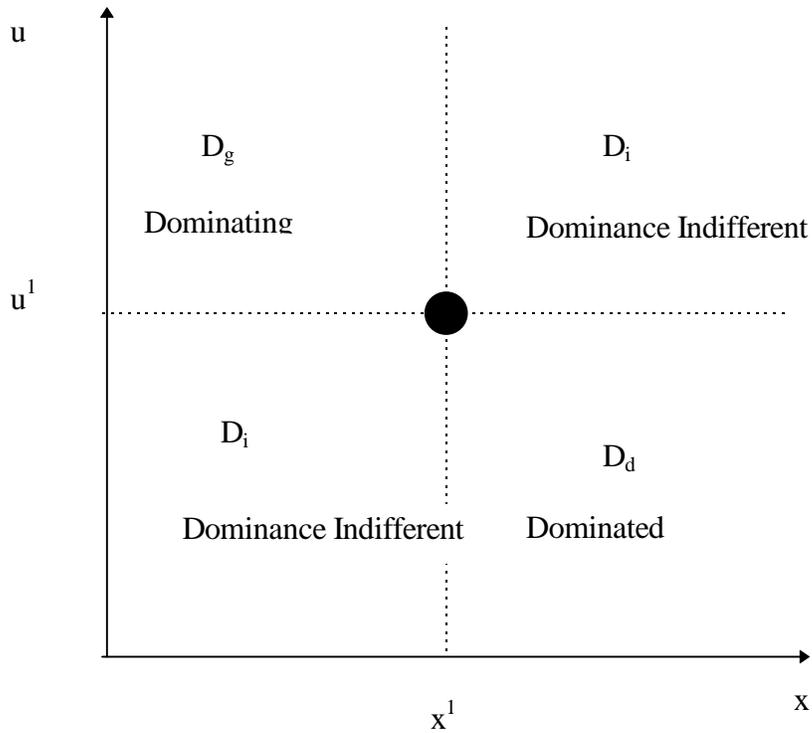
Figure 2-2. One Input and One Output Production System
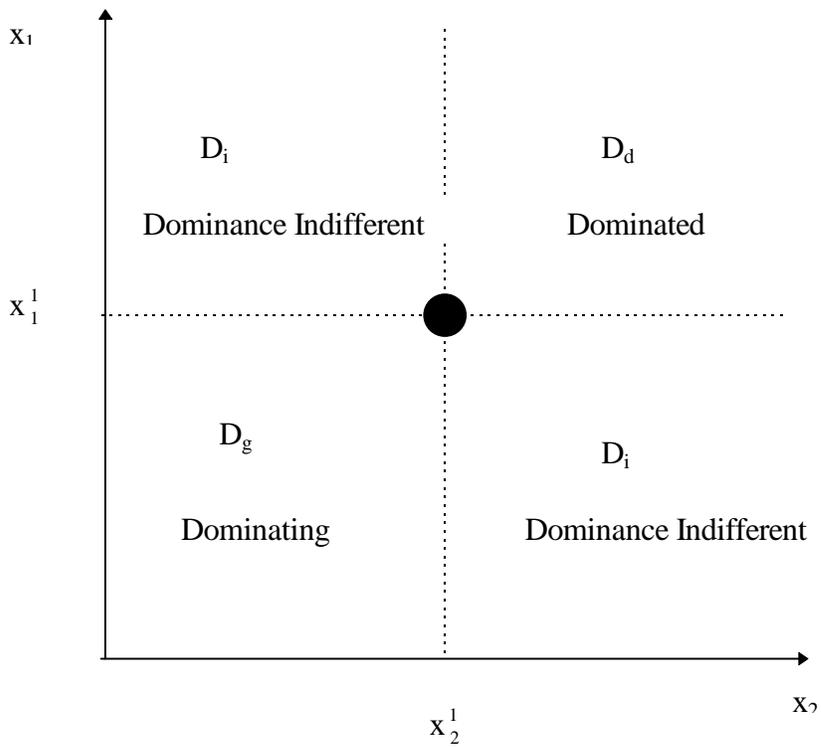
Figure 2-3. Two Input Production System (output held constant)
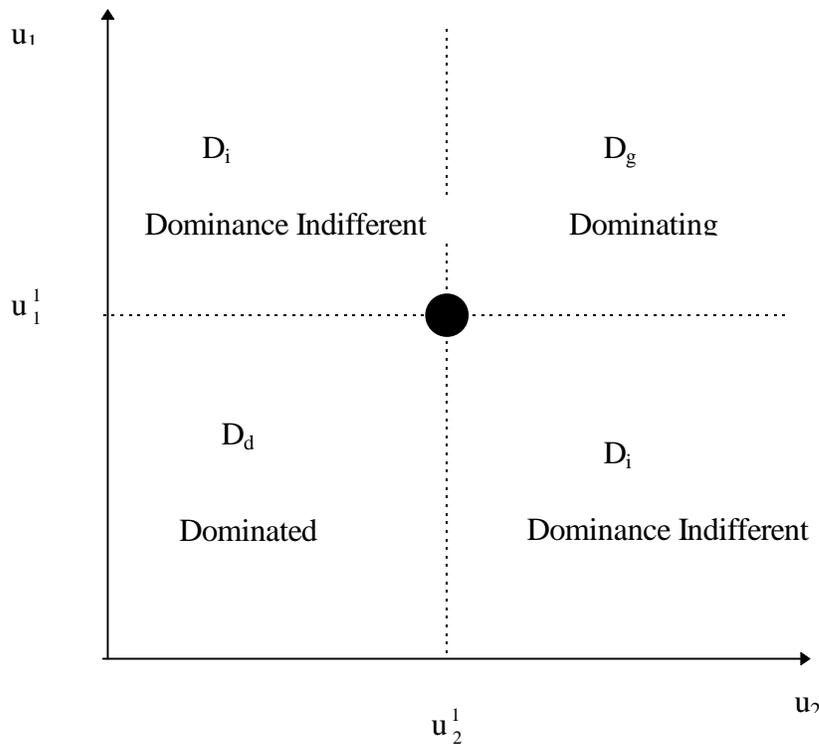
Figure 2-4. Two Output Production System (input held constant)

The efficiency of a particular observation at a particular time is defined as a performance index. Performance indexes exist for both progress and regress and for both inputs and outputs. The definitions progress and regress are shown below (Tulkens and Vanden Eeckaut, 1995, p. 484). These definitions are for panel data of k production systems. For simple time series data for a single production system the superscript k is eliminated from the definiton.

Definition 1: An observation k (x^{kt}, u^{kt}) is said to induce 'progress' if it is (i) undominated at time t, and (ii) dominating one of several observations, made at some time $s < t$ and found undominated at s.

Definition 2: An observation k (x^{kt}, u^{kt}) is said to induce 'regress' if it is (i) undominated at time t, and (ii) dominated by one or several observations, made at times $s < t$ and found undominated at s.

Performance indexes in inputs and outputs are defined to be (Tulkens and Vanden Eeckaut, 1991, p. 13):

For regress observation in inputs:

$$R^i(t) = \underset{d \in D_{0B(t)}^{g}}{Min} \left\{ \underset{i=1,\ldots I}{Max} \left(\frac{x_i^d}{x_i^t} \right) \right\} \qquad \text{(Eq. 2-29)}$$

For regress observation in outputs:

$$R^o(t) = \underset{d \in D_{0B(t)}^{g}}{Min} \left\{ \underset{j=1,\ldots J}{Max} \left(\frac{u_j^t}{u_j^d} \right) \right\} \qquad \text{(Eq. 2-30)}$$

For progress observations in inputs:

$$P^i(t) = \underset{d \in D_{0B(t)}^{d}}{Max} \left\{ \underset{i=1,\ldots I}{Min} \left(\frac{x_i^d}{x_i^t} \right) \right\} \qquad \text{(Eq. 2-31)}$$

For progress observation in outputs:

$$P^o(t) = \underset{d \in D_{0B(t)}^{d}}{Max} \left\{ \underset{j=1,\ldots J}{Min} \left(\frac{u_i^t}{u_j^d} \right) \right\} \qquad \text{(Eq. 2-32)}$$

The regress measure in inputs (Eq. 2-29) first takes the minimum ratio of the inputs of two observations. One observation is the one that is being measured (denominator), and the other observation is from the set of benchmark observations (numerator). This minimum ratio is then taken for each of the observations that are part of the benchmark correspondence set with the maximum of all of these ratios being the metric of regress for inputs. So, this metric is selecting the input for which the difference is the least between the regress observation and then taking the maximum of all of these minimum ratios. This will tend to produce a more stable metric that does not depend on outlier observations. The other regress and progress metrics are similarly defined. The more intuitive meaning of this in

terms of a measure of performance is not clear from the literature. As will be discussed in Chapter 3 this metric is not used because of the difficulty in interpreting its meaning.

2.2.3.2 *Free Disposal Hull*

FDH differs from the Benchmark Correspondence method is that the Dominating production plans are used as the reference set (in Benchmark Correspondence the production plans that are neither dominant or dominating are the reference set). Efficiency is then measured as the distance of a particular production plan to the dominating production plans (frontier). The Free Disposal Hull method differs from other Data Envelopment Analysis method in that the requirement of convexity is relaxed. The FDH method was introduced by Deprins, Simar, and Tulkens (1984) as a more practical approach for measuring efficiency based on actual observations rather than an assumed production function.

The observations that define the frontier for FDH are the dominating points in the set, D_g, defined above as part of the definition of the benchmark correspondence method. For a given data set:

$$Y_0 : \ Y_{FDH} \subset Y_{DEA\text{-}ID} \subset Y_{DEA\text{-}CD} \subset Y_{DEA\text{-}C} \qquad \text{(Eq. 2-33)}$$

Where the DEA formulations are Variable Returns to Scale (termed ID for Increasing only in the lower range of the inputs up to some point, and Decreasing beyond), CD (meaning constant returns to scale up to a certain point and then decreasing returns to scale), and C frontier points.

In an application to data from urban transit, retail banking, and courts, a comparison is made between FDH and DEA (using different postulates as defined above) Tulkens (1993). This comparison of methods shows that FDH measures fit the data better. As a result efficiency scores are always higher than DEA methods that have a more restrictive definition of the frontier.

44

FDH as applied to time series data is discussed by Thiry and Tulkens (1990). Efficiency is evaluated sequentially only for observations $k \leq t$. At each successive time, t, the new observation is evaluated to determine whether or not it is part of the FDH frontier. The observation at time, t, is sequentially inefficient if it is not part of the FDH frontier (dominated) or is sequentially efficient is it is part of the FDH frontier (undominated). There is another possible outcome. An observation may be declared sequentially efficient and dominant observations previously declared efficient. This indicates local technical progress.

De Borger and Kerstens (1996) calculate efficiency for Belgian local governments with FDH using radial and nonradial measures of efficiency. Radial in this paper refers to a radius from the origin through the observation being measured to the frontier as defined by Debreu (1951) and Farrell (1957). Other non-radial measures are applied. The analysis shows "that the differences in ranking between radial and nonradial measures were more pronounced than those among the non-radial alternatives" (De Borger and Kerstens, p. 56, 1996).

Bardham, Bowlin, Cooper, and Sueyoshi (1996) note that an efficient production plan (or DMU) for the FDH method may have slacks. The FDH linear program is reformulated to require that all slacks be zero for a production plan to be efficient. In addition, the formulation restricts the measurement of efficiency of a production plan to a distance to another actual production plan which has been determined to be efficient. A production plan is efficient only if there is no other production plan which dominates that production plan. Of the efficient production plans, only the most efficient is selected as the basis for measuring productive efficiency of a non-efficient production plan.

2.2.3.3 *Other Dominance Perspectives*

Hougaard and Tvede (1997) view DEA and FDH as special cases of dominance analysis. They propose a performance vector that is simply the ratios of the inputs and

outputs of two points (Hougaard and Tvede, p. 8, 1997). Let $c = (x_c, u_c)$ and $d=(x_d,u_d)$. The vectorial degree of dominance for c relative to d is defined as:

$$V(c,d) = \left(\left(\frac{x_{di}}{x_{ci}} \right)^I_{i=1}, \left(\frac{u_{cj}}{u_{dj}} \right)^J_{j=1} \right). \qquad \text{(Eq. 2-34)}$$

It is noted that V is independent of the units of measurement (i.e. commensurability) and has strict monotonicity and homogeneity. Hougaard and Tvede (1997, p. 13) suggest the use of tables to compare performance evaluations over time to determine the extent to which a particular observation is ahead or behind the contemporaneous reference set.

Tulkens (1993a) explores a number of options for measuring performance based on dominance. For measuring the efficiency dominance of a particular production plan options explored are comparison to another production plan; comparison to the set of all production plans; and comparison to the set of observations itself. For comparison to another production plan Tulkens introduces an average metric. For an input index where x^k is dominated by x^d and there are I inputs the average metric is:

$$PD^i_1 (k;d) = \left[\frac{\sum_{i \in I} x^d_i / x^k_i}{I} \right] \le 1 \qquad \text{(Eq. 2-35)}$$

For comparisons of a production plan to the set of all production plans there are integer measures, extreme measures, and average measures. Integer measures are simply counts of the number of production plans in the dominating set, dominant set, and dominance indifferent sets relative to the production plan being evaluated. Extreme measures take the production plan that dominates the most or is dominated the most. So called Full Measures are introduced by Tulkens (1993a). These measures use the average index and the other metrics discussed above as the numerator and use integer measures, such as the count of the number of dominated production plans, as the denominator of the metric. Fried, Knox Lovell, Vanden Eeckaut (1993) provide a total count of dominated, undominated, and dominance indifferent relationships as part of an analysis of the productive efficiency of US credit unions.

2.3   Related Topics

This sections reviews topics that are related to the measurement of productive efficiency.   Multiobjective programming, reviewed in Section 2.3.1, is relevant since environmental objectives and productivity objectives may differ.   Some method of combining multiple objectives is required. Input substitution, reviewed in Section 2.3.2, relates the trade-off and the relative desirability of different inputs. Some of the approaches that have been taken to measure environmental performance are then discussed in Section 2.3.3. The related topic of cost accounting is then reviewed in Section 2.3.4. Then, data analysis techniques that are relevant to the reduction and evaluation of production data are discussed in Section 2.3.5.   The related issue of aggregation of data is then discussed in Section 2.3.6. Finally, the issue of the evaluation of objectives that are expressed in terms of ordinal classification is discussed.

2.3.1 *Multiobjective Linear Programming*

Simon (1948) describes the decision making environment of an organization as one where "decision making processes are aimed at finding courses of action that are feasible or satisfactory in the light of multiple goals and constraints..." (Simon, 1948, p. 274). This basic idea that organizations in fact have sometimes competing objectives making an optimal solution impractical is the basis of multi-objective programming. This research addresses the conflicting objectives of increasing efficiency and reducing environmental impact. This research hypothesizes that there is some correlation between these two objectives, making it possible to increase efficiency and reduce environmental impact. However, some environmental goals may only be achievable by reducing efficiency as measured with standard approaches. Multiobjective linear programming is one approach that may be applied to evaluate these potentially conflicting objectives of increased productive efficiency and reduced environmental impact.

Multi-objective objective methods have been applied to a large variety of problems with goal-programming being the most prevalent approach in published studies (White, 1990). The objective function for a multi-objective problem may be represented as:

Maximize $f(x) = [f_1(x), f_2(x),...,fp(x)]$ subject to $x \in X$. (Eq. 2-37)

where,

p = number of objective functions.

A solution that maximizes each of the objective functions simultaneously is called a Superior solution. Since at least some of the objective functions are typically conflicting a superior solution is usually not obtained. Instead, "an efficiency solution is one for which there does not exist another feasible solution which does at least as well on every single objective, and better on at least one objective" (Evans, 1984, p.1270).

There are three basic approaches to multiple-objective techniques. These are (Ignizio, 1982, p. 374-375).

1. Weighting or utility methods: Dollars, utility methods, and other approaches are used to derive weights for objectives that are then used to produce a single measure. Solution then proceeds using single-objective methods.

2. Ranking or prioritizing methods: Objectives are ranked according to their perceived importance.

3. Efficient solution (or generating) methods: The total set of efficient solutions are generated.

Each of these methods has potential drawbacks for the research being proposed. The assignment of weights requires subjective consensus judgments. Previous studies of the assignment of weights have shown that "[n]ormative theory is not found totally supported by empirical behavior" (Olson, 1992, p. 17). Solutions generated using ranking methods are difficult to relate to the satisfaction of the ranked objectives (Ignizio, 1982, p. 375). The number of possible solutions may be too large to evaluate efficiently (Evans, 1984, p. 1273).

One of the most used approaches for multi-objective programming is goal programming. Goal programming seeks to "achieve a set of desired but adjustable goals as

48

closely as possible by minimization or penalizing deviations from the goals" (Min, 1991, p. 301). This is expressed mathematically as (Min, 1991, p. 301):

$$\text{minimize} \sum_{i=1}^{m} W_i^+ d_i^+ + W_i^- d_i^- \qquad \text{(Eq. 2-38)}$$

subject to

$$\sum_{j=1}^{n} a_{ij}x_j - d_i^+ + d_i^- = g_i \quad \text{for all i}$$

$$x_j,\ d_i^+,\ d_i^-,\ W_i^+,\ W_i^- \ge 0 \ \text{ for all i, j}$$

where,

x_j = decision variable;

a_{ij} = technological coefficients;

g_i = goals;

d_i^+ = respective deviation from the goals

$$= \frac{1}{2}\left[\left|\sum_{j=1}^{n} a_{ij}x_j - g_i\right| + \left(\sum_{j=1}^{n} a_{ij}x_j - g_i\right)\right];$$

d_i^- = respective deviation from the goals

$$= \frac{1}{2}\left[\left|\sum_{j=1}^{n} a_{ij}x_j - g_i\right| - \left(\sum_{j=1}^{n} a_{ij}x_j - g_i\right)\right];$$

W_i^+, W_i^- = preassigned weights representing relative importance of deviations.

There are many approaches for the assignment of weights and the solution of the generalized linear program presented above. Preemtive priorities may also be assigned where a priority factor is assigned to each of the goals. This priority is then used as part of a modified simplex solution method to consider lower importance goals only after higher importance goals are achieved (Canada and Sullivan, 1989. p. 289).

A variation on goal programming is goal interval programming. Rather than the minimum value being achieved when x = g, the goal is expressed as an interval so that the minimum value is reached when $g_1 \leq x \leq g_2$ (Charnes and Cooper, 1977). Another variation is to add additional preemtive weights to the formulation that are interpreted to mean that no substitutions can occur across categories (Charnes and Cooper, 1977).

Romero (1985) notes that goal-programming can be thought of a special case of the following distance function model:

$$\text{Min} \quad \left[\sum_{j=1}^{n} W_j \left| \hat{f}_j - f_j(x) \right|^P \right]^{1/P} \tag{Eq. 2-39}$$

$$\text{s.t. } x \in F$$

where,

W_j = weight attached to the objective j;

$\hat{f}_j$ = aspiration level for the objective j;

$f_j(x)$ = objective function j

F = feasible set.

P = is some constant

Goal programming has been applied in Data Envelopment Analysis. In particular goal programming has been applied to DEA based analyses of stochastic efficiency. A goal programming method for allocative data envelopment analysis (ADEA) is developed by Retzlaff-Roberts and Morey (1993). The objective function is composed of a two sided noise component and a term representing inefficiency with a coefficient. The coefficient of the inefficiency term that must be specified as part of the model formulation and is equivalent to a goal.

Charnes, Cooper, et. al. (1976) applied goal interval programming to aid resource allocation decisions in the U.S. Coast Guard's Marine Environmental Protection Program. Goals in this formulation were the various operational requirements of the Coast Guard enforcement program to prevent discharges of oil and other hazardous material into the

marine environment. It is noted that "[f]or full exploitation as a management aid, however, this model should be developed subsequently as an iterative and interactive technique with goal alterations and functional weights examined in a variety of resource availability contexts" (Charnes, Cooper, et. al., 1976, p. 360). In an application to waste management, waste are generated and then transformed by treatment processes that then produce residual wastes that are disposed of to the environment (Ponagiotakopulos, 1975). An example is presented where goals include air and water discharge limits, population level, industrial activity, and cost of the waste management system.

In a review of weight restrictions used in DEA (Allen, Athanassopoulos, Dyson, and Thanassoulis, 1997), the approaches closely parallel those taken in multi-objective programming in general. There are basically three approaches identified by Allen, et. al. to be: direct restriction on the weights, adjusting the observed input-output levels, and restricting the virtual inputs and outputs. Direct restrictions on weights place various restrictions on the values of μ_j and ν_i (variables in the linear programming DEA model). Restrictions on inputs and outputs have either been a variation on assurance region restrictions or have imposed an ordinal relationship on the DEA weights. Restricting virtual inputs and outputs is done by restricting the proportion of virtual output (or input) of DMU_k. Since these restrictions are DMU specific it can be difficult to apply. Belton and Vickers (1993) replace DEA with a multiple criteria procedure that is implemented on a computer program that allows decision makers to assign weighs to inputs and outputs (rather that allowing the DEA procedure to assign these weights). An interactive computer program is used to allow decisions makers to more easily interpret results.

2.3.2 *Input Substitution*

Production theory, where a specific functional form of a production function is assumed and all firms are assumed to be minimizing costs or maximizing profit, allows the effects of substitution among inputs to be exactly specified. The Rate of Technical

Substitution (RTS) between two inputs where q is the output is defined to be (Henderson and Quandt, 1980, p. 70):

$$\text{RTS} = \frac{dx_2}{dx_1} = \frac{\partial q / \partial x_1}{\partial q / \partial x_2} \qquad \text{(Eq. 2-40)}$$

Another measure often used is the Elasticity of Substitution which "is a pure number that measures the rate at which substitution takes place. It is defined as the proportionate rate of change of the input ratio divided by the proportionate rate of change of the RTS" (Henderson and Quandt, 1980, p. 73).

$$\sigma = \frac{d \ln(x_2 / x_1)}{d \ln(f_1 / f_2)} = \frac{f_1}{f_2} \frac{d(x_2 / x_1)}{d(f_1 / f_2)} \qquad \text{(Eq. 2-41)}$$

where:

$f_1 = \partial q / \partial x_1$ and

$f_2 = \partial q / \partial x_2$

Applications of production theory to the measurement of factor substitution follow the general pattern of assuming a particular form of the production function and then using aggregated data to investigate substitution of inputs. For example, Humphrey and Moroney (1975) examined the hypothesis that capital equipment, labor, and natural resources are significantly substitutable as productive inputs. Using the transcendental logarithmic production function (i.e., translog), and data from U.S. Bureau of the Census data and Office of Business Economics, inputs are estimated for manufacturing industries such as electrical machinery and primary metals. This high level analysis reveals that "within most of these resource-intensive sectors the substitutability between capital and resource products as well as between labor and resource products appears to be at least as high as the substitutability between capital and labor" (Humphrey and Moroney, 1975, p. 70). It should be noted that there are different methods of calculating the elasticity of substitution that are generally not equivalent (Segerson, 1989).

The use of data envelopment methods does not allow the straightforward calculation of technical substitution or elasticity of substitution. Some work where actual data is used

to estimate rates of technical substitution rather than an assumed production function has been done. Kopp and Smith (1978) use engineering, process analysis models of iron and steel plants to evaluate the performance of neoclassical cost function in estimating input substitutions. Output from the models are used to construct a translog cost function. This approximation of the technology is then compared to the results from the model of the cost function of the production technology. Capital inputs are divided into productive plant and equipment and pollution abatement equipment. Other inputs are iron ore, fuel oil, natural gas, scrap, labor, maintenance and all other operating inputs. Constraints on emissions are considered in the simulation models of the production process. Since the simulation model produces discrete output, substitution effects are measured with a "derived elasticities of substitution" defined to be (Kopp and Smith, 1978, p. 637):

$$\sigma_{ij}^{A} = \frac{1}{C_{2j}} \frac{\left[\dfrac{Q_{2i} - Q_{1i}}{Q_{2i} + Q_{1i}}\right]}{\left[\dfrac{P_{2j} - P_{1j}}{P_{2j} + P_{1j}}\right]} \qquad\qquad \text{(Eq. 2-42)}$$

where,

C_{2j} = the cost of the jth factor input at P_{2j} prices;

Q_{2i} = the quantity of the ith factor input evaluated at P_{2j} for the jth factor price where all other prices and output held constant at a base level; and

Q_{1i} = the quantity of the ith factor evaluated at a price P_{1j} for the jth factor and all other prices and output held constant at a base level.

This analysis reveals that progressive aggregation of inputs tends to reduce the neoclassical models ability to identify input relationships. The analysis also appears to show that the neoclassical model does not adequately capture the effects from the production of intermediate inputs produced by the plants as byproducts of treatment processes. Finally, "results suggest that the presence of discharge constraints together with the failure to reflect them in the specification of the cost function reduces the neoclassical model's ability to identify input relationships" (Kopp and Smith, 1978, p. 650).

2.3.3 *Environmental Performance Measurement*

Performance measurements are intended to provide information to managers and workers to allow then to improve performance. The measurement of environmental performance poses important challenges that are summarized by Well, Calkins, and Balikov (1994, p. 53) to be:

- Success measures must address multiple objectives in a balanced manner.

- Value judgments must be made, but it must be possible to interpret the effect of these value judgments on results.

- We cannot eliminate scientific uncertainty; we can only address it directly, making scientific judgments transparent and providing flexibility to test alternative assumptions.

- Effective environmental management systems must be tied into corporate management information systems to minimize data gathering and data manipulation burdens.

- A measurement system should provide a broad framework to track and categorize performance, but should retain flexibility to adapt to different types of operations and technological changes.

As noted by Rooney (1992, p. 258) "[w]ithout simple, timely control reports, waste soon returns to its original level."

In is increasingly being recognized that there is a relationship between environmental performance and overall economic performance. The Electronics Industry Environmental Roadmap (1994) asserts that the global competitiveness of the companies of the electronics industry will depend upon the ability to:

- Minimize waste.
- Enhance productivity.
- Minimize material costs.
- Reduce hazardous materials usage.
- Reduce energy consumption.
- Re-engineer facilities processes, products, and management systems.

The next sections provide background on performance measures for pollution prevention. First, classifications that may be applied to inputs and outputs are reviewed. Then, relevant examples of metrics are reviewed.

2.3.3.1 *Input and Output Classification*

Any metric for pollution prevention and environmental performance will have to measure reductions in inputs and outputs. Various distinctions have been made on the classification of inputs and outputs for the purposes of reducing and measuring environmental impacts. This section reviews some of the classifications developed for inputs and outputs.

Kraft (1992) classifies process streams into four categories for waste minimization analysis. These are nonuseful (waste), feed, intermediate, and recycle. Lave, Hendrickson, and McMichael (1994, p. 19A) list some ways to handle "products that no longer have a use:"

1. discarded into the environment;
2. placed in a permitted landfill;
3. burned within a permitted waste-to-energy incinerator, producing electricity;
4. put to a low-value use, sometimes after breaking it down into its components (e.g., polystyrene cups can be made into part benches, 727 aircraft can be used as reefs);
5. put to high-value use (e.g., aluminum cans can be melted to make new ones);
6. rebuild, with some components discarded, and reused (e.g., automobile water pumps);
7. reused (e.g., returnable, refillable beverage bottles).

Categories 1 and 2 are not recycling; categories 5 to 7 are regarded as recycling. Categories 3 and 4 are ambiguous: they don't create litter, but they don't use materials for a high-value product. Lave, Hendrickson, and McMichael (1994) go on to discuss how the above one-

dimensional view of recycling is too simplistic. In addition to reductions in virgin material, recycling decisions should also consider energy use, environmental loading, labor, and product quality. Strategies for reducing virgin materials use such as increasing the economic life of products are also not captured.

The product life components are defined by Keoleian, et. al. (1994a, p. 16-17) as: product, process, distribution, and management. There are various ways in which inputs and outputs may be classified from this life cycle perspective which are shown in Tables 2-5 through 2-10. These tables use the term residual to mean waste products. When some waste is treated there will always be some by-product of this treatment (e.g., sludge, heat, filter cake, and so on) that is a waste and is often referred to as a residual.

Table 2-5. Classification of Inputs and Outputs by Life Cycle Stage
(Keoleian, Menery, Vigon, et al., 1994)

Product	*Inputs*
	Materials
	Energy (embodied)
	Outputs
	Products, co-products, & residuals
Process	*Inputs*
	Materials
	direct: process materials
	indirect: 1st level (equipment & facilities)
	2nd level (capital & resources to produce 1st level)
	Energy: process energy (direct & indirect)
	People (labor)
	Outputs
	Materials (residuals)
	Energy
Distribution	*Inputs*
	Materials
	packaging
	transportation
	direct (e.g., oil & brake fluid)
	indirect(e.g., vehicles and garages)
	Energy
	packaging (embodied)

```
                            transportation (Btu/ton-mile)
                   People (labor)
            Outputs
                   Materials(residuals)
Management  Inputs
                   Materials, office supplies, equipment & facilities
                   Energy
                   People
                   Information
            Outputs
                   Information
                   Residuals
```

Table 2-6. Classification of Inputs and Outputs by Environmental Impact
(Keoleian, Menery, Vigon, et al., 1994)

```
                            Materials

Amount (Intensiveness)

Type
   Direct
            product related
            process related
   Indirect
            fixed capital (bldg. & equipment)

Source
   Renewable
            forestry
            fishery
            agriculture
   Nonrenewable
            metals
            nonmetals
Character
   Virgin
   Recovered (Recycled)
   Reusable/Recyclable
```

```
        Useful Life
        Resource Base factors
                location
                        locally available
                        regionally available
                scarcity
                        threatened species
                        reserve base
                quality
                        composition
                        concentration
                management/restoration practices
                        sustainability


Impacts associated with extraction, processing, and use
        Residuals
        Energy
        Ecological Factors
        Health and safety
```

Table 2-7. Classification of Energy Sources by Environmental Impact
(Keoleian, Menery, Vigon, et al., 1994)

```
                            Energy

Amount (energy efficiency)

Type
    Purchases
    Process by-product
    Embodied in materials

Source
    Renewable
            wind
            solar
            hydro
            geothermal
            biomass
```

```
        Nonrenewable
                fossil fuel
                nuclear


Character
    Resource base factors
                location
                scarcity
                quality
                management/restoration practices


Impacts associated with extraction, processing, and use
    Materials
    Residuals
    Ecological factors
    Health and safety
    Net energy
```

Table 2-8. Classification of Residuals by Environmental Impact
(Keoleian, Menery, Vigon, et al., 1994)

```
                              Residuals

Type
    Solid waste
                solid
                semi-solid
                liquid
    Air emissions
                gas
                aerosol
                particulate
    Waterborne
                dissolved
                suspended solid
                emulsified
                chemical
                biological


Characterization
    Nonhazardous
```

```
                constituents
                amount
        Hazardous
                constituents
                toxicity
                concentration
                amount
        Radioactive
                potency/half life
                amount
                concentration

        Environmental fate
                Containment
                Degradability (physical, biological, chemical)
                Bioaccumulation
                Mobility/Transport mechanisms
                        atmospheric
                        surface water
                        subsurface/groundwater
                        biological
                Treatment/Disposal
                        impacts
                                residuals
                                energy
                                materials
                                health & safety effects
```

Table 2-9. Ecological Factors
(Keoleian, Menery, Vigon, et al., 1994)

```
                        Ecological Factors


Type of ecosystems impacts
        Physical (disruption of habitat)
        Biological
        Chemical

Ecological stressors
        Diversity
        Sustainability
```

Rarity
Sensitive species
Scale
Local
Regional
Global

Table 2-10. Health and Safety Factors
(Keoleian, Menery, Vigon, et al., 1994)
Human Health and Safety

Population at risk
    Workers
    Users
    Community

Toxicological characterization
    Morbidity
    Mortality
    Exposure
        routes
            inhalation
            skin contact
            ingestion
        duration
        frequency

Nuisance effects
    Odors
    Noise
Accidents
    Type

2.3.3.2 *Metrics*

Measurement is essential to the integration of environmental issues into business practices. Some of the metrics listed in the Electronics Industry Environmental Roadmap (1995, p. 21) are:

- Energy metrics:

    - Total energy consumed during the product life-cycle.
    - Renewable energy consumed during the product life cycle.
    - Power used during operation (for electrical products).

- Emission metrics:

    - Toxic or hazardous materials used in production.
    - Total industrial waste generated during production.
    - Hazardous waste generated during production or use.
    - Air emissions and water effluents during production.
    - Greenhouse gases and ozone-depleting substances released.

- Materials management metrics:

    - Useful operating life.
    - Product disassembly and recovery time.
    - Percentage of recycled materials used as input to product.
    - Percentage of recycled materials available at end-of-life.
    - Percentage of product disposed or incinerated.
    - Percentage of packaging or containers recycled.

- Economic metrics:

    - Average life cycle cost incurred by the manufacturer.
    - Average life cycle cost incurred by the manufacturer.
    - Purchase and operating cost incurred by customers.
    - Cost savings associated with design improvements.

Keoleian (1994, p. 46) states that environmental requirements should minimize:

    - raw materials consumption
    - energy consumption
    - waste generation
    - health and safety risk
    - ecological degradation

In a study of pollution prevention programs in Ohio, Butler (1996) summarized pollution prevention measurement methods to be:

- Actual Quantity Change (AQC) - the change in annual quantity of waste generated. This assumes that a decrease in waste generated at a facility represents pollution prevention.

- Adjusted Quantity Change (ADQC) - adjusted quantity measures account for changes in production, or other business activity by an index (also knows as a activity or production index.) This assumes that waste generation and production levels are linearly related.

- Materials Accounting - quantitatively tacks materials through a production process where pollution prevention activity is likely to occur. "Materials accounting information can be manipulated to develop facility and process specific efficiency ratios. A lower ratio indicates that a smaller portion of a processes or facility's raw materials are wasted and a higher portion or the raw materials are going to productive use" (Butler, 1996, p. 20).

- Pollution Prevention Frontiers - based on the principles of Materials Accounting - utilizes an index of pollution prevention (waste generation per unit of production) to measure pollution prevention effectiveness.

- Toxicity Change - assessing the increase or decrease of a waste's toxicity or hazard level is another important avenue which may indicated pollution prevention.

Butler (1996) evaluated the AQC and ADQC methods using environmental compliance data from companies in Ohio. It was concluded that "[t]he application of AQC and ADQC measurements methods were largely ineffective in making definitive pollution prevention conclusions" (Butler, 1996, p. 45). The other three methods were not evaluated.

Rooney suggests the Absolute Material Utilization Ratio (AMU%) as such a measure. The "AMU% is the ratio of material which is absolutely needed in the good output to the material actually used in production" (Rooney, p. 258). "Watanabe proposes a Resource Productivity measure for evaluating "industrial performance compatible with environmental preservation." The resource productivity is defined as (Keoleian and Menerey, 1994): [(Economic value added) * (Product Lifetime)]/[(Material consumed-recycled) + (Energy consumed for production, recycled) + (Lifetime energy used)]. Legarth (p. 127, 1995) proposes a Resource Recovery Efficiency (RRE) and a Resource Recovery Potential (RRP)

metric to gauge resource recovery. "The RRE states how much of the RRP is realized by a recycling system" (Legarth, 1995, p. 127). RRP is "a measure of how long the use of the primary resource may be prolonged because of the recycling action - using today's primary consumption rate as the basis of the prognosis. The RPE divided by the RP can be interpreted as efficiency. These measures are: (Legarth, 1995, p. 128):

$$\text{RRP} = \sum_i \frac{F_i}{P_i} \cdot \frac{C_i}{R_i} \quad \approx \sum_i \frac{F_i}{R_i} \qquad \text{(Eq. 2-43)}$$

$$\text{RRE} = \sum_i E_i \cdot \frac{F_i}{P_i} \cdot \frac{C_i}{R_i} \quad \approx \sum_i E_i \cdot \frac{F_i}{R_i} \qquad \text{(Eq. 2-44)}$$

where,
F is the amount of resource in one ton of scrap
P is the annual primary production of resource
C is the annual consumption of primary resource
R is the world reserves of the resource
I is the number of resources in scrap
E is the percentage of the resource I actually recovered by the recycling system

The 'complexity' of a product also has an environmental impact. More complex products tend to produce more scrap for example. Cooper, Sinha, and Sullivan (1992) define the complexity of a system "as the number of elements and the number of first-order interrelations between elements in the system to be considered" (Cooper, Sinha, Sullivan, 1992, p. 42).

2.3.4   *Cost Accounting*

Inputs such as capital, labor, overhead, and the cost of material inputs and outputs are obtained from the cost accounting system. In the case of capital and direct labor, data from a typical cost accounting system can be used directly in the analysis of performance. Other data obtained from a cost accounting system are dependent on assumptions that are part of the cost accounting system as well as how data is collected.

One definition of the purpose of the cost accounting system in a manufacturing environment is: "Accounting provides routine cost and performance data to management by setting standard costs against which actual data are compared in order to measure cost performance of a production shop" (Koons, 1990, p. 9). The traditional approach to cost accounting is to assign overhead expenses to a product based on labor content. This approach has many well-recognized defects. Drucker (1990) identified four limitations of traditional cost accounting:

1. Cost accounting is based on "the realities of the 1920's" when direct labor was 80% of manufacturing costs other than raw materials.

2. The benefits of process changes are primarily defined in terms of labor cost savings.

3. Only the cost of producing is measured, the cost of non-producing (e.g., machine down time, quality defects) is ignored.

4. A factory is treated as an isolated entity where costs and benefits such as a products acceptance in the marketplace or service quality are not considered.

The practical effect of these defects is that cost accounting systems do not provide the information required to manage manufacturing operations. In a survey of 32 Canadian manufacturing companies the following issues emerged: piecemeal selection of equipment [i.e., machine capacities not balanced to production demand], costing work-in-progress, rush orders, marketing and manufacturing coordination, set-up costs, and downscaling" (Mackey, p. 35).

At a more practical level "compacting all overhead expense information into a single burden rate makes it difficult for engineers to do analysis and managers to make decisions. As a result, many engineering studies are done apart from the accounting database, causing the generation of inconsistent results..." (Koon, p. 11). The Hewlett-Packard Roseville Networks Division found this to be the case and has moved to Activity Based Accounting with the goal of "accurately reflect manufacturing costs, use data that manufacturing could

collect easily, and meet the legal and practical needs of the accounting function" (Berlant, 1990, p. 179). A simple example provided by Koons (1990, p. 12) shows "labor expense is allocated by labor hours, facility expense is allocated by machine time, and support expense by volume."

The assignment of overheads to a particular product or product line can vary substantially based on the assumptions made by the cost accounting system. In a study of overhead costs by Foster (1990) the mean of the components of manufacturing cost across 37 facilities in the electronics industry were found to be: Direct Materials (54.3% of total), Direct Labor (6.6% of total), and Manufacturing Overhead (39.1% of total). Activity Based Accounting is one approach for overcoming the problems inherent in traditional cost accounting systems.

Activity based cost accounting provides information about the work or activity that consumes resources and delivers value in a business. Such information allows managers to focus on the actual sources of cost of profit. Johnson (1988) categorizes activity based information into two types. One type is non-financial and includes information about sources of competitive advantage such as quality, flexibility and cost. The second type is strategic cost information that enables managers to assess the long-term profitability of a company's mix of products and activities. This system allows activities that are not generating value, such as scrap production and overtime for end-of-period production spurts, to be identified. Johnson's four steps for managing waste are: "chart the flow of activities throughout the organization; identify sources of customer value in every activity, and eliminate any activities that contribute no identifiable value to customers; identify causes of delay, excess, and unevenness in all activities; and track indicators of waste" (Johnson, 1988, p. 90).

In an Environmental Management System cost accounting is a key element. Life-Cycle Costing and Activity Based Accounting may be applied to allocate environmental expenditures to products. There are four elements of environmental costs. These are the

"Usual Costs and Operating Expenses" such as equipment, labor, materials, and so on. There are "Hidden Regulatory Costs" such as reporting, permitting, monitoring, testing, training, and inspection. There are "Contingent Liability Costs" that can include penalties and fined, legal claims, and awards. Finally, there are "Less Tangible Costs" that could include corporate image, and not-meeting consumer demand for environmentally friendly products (Krueze, p. 39, 1994).

From the perspective of implementing an Environmental Management System there are a number of problems that exist with accounting systems as identified by Willits (1994). One is a lack of accounting for materials costs once a material is released for use. As a result the cost of material inputs going to waste is not considered. Further, the cost of waste disposal is not considered to be a part of the cost of materials which leads to inaccurate costing of repairs and maintenance activities. Willits (1994) suggest that the financial accounting system be coupled with the environmental management system to track materials from purchase to disposal. This allows more assignment of environmental costs to materials rather than burying these costs in overhead as is often the case now.

One of the criticisms of traditional cost accounting systems is the assumption that allocated costs assigned by labor hours or machine hours or some other bases are directly proportional to volume of product items manufactured. There are many costs, however, that vary with the diversity and complexity of products. This cost of complexity is summarized by (Cooper and Kaplan, 1988, p. 121):

> [S]upport-department costs vary not with the volume of product items produced rather they vary with the range of items produced (i.e., the complexity of the production process). The traditional definition of variable cost, with its monthly or quarterly perspective, views such costs as fixed because complexity-related costs do not vary significantly in such a short time frame. Across an extended period of time, however, the increasing complexity of the production process places additional demands on support departments, and their costs eventually and inevitably rise.

One solution to this problem is to assign costs based on transactions such as the number of setups that reflect the complexity of a manufacturing operation. Costs may also be assigned based on the total number of different parts used to reflect the increasing complexity that occurs with large numbers of unique parts for low volume products.

2.3.5   *Data Analysis*

As noted by Seaver and Traintis "[g]iven that most data used for production studies has not been accumulated for such purposes, it is important that data and modeling issues such as, outliers, collinearity, measurement errors, aggregation, input/output specification and others are studied carefully given their impact on efficiency performance measurement" (p.1, 1991). This section reviews data analysis issues and techniques relevant to the research. Outlier analysis is first discussed, followed by a discussion of cluster analysis and the normalization of data.

An outlier in a set of data is defined by Barnett and Lewis (1994, p. 7) to be "an observation (or set of observations) which appears to be inconsistent with the remainder of that set of data." Variability that gives rise to outliers may come from one of three sources summarized in Barnett and Lewis (1994, p. 33): Inherent Variability (natural feature of the population), Measurement Error, and Execution Error (imperfect collection of data - e.g., choosing a biased sample). There are two basic approaches for dealing with outliers: rejection or accommodation. Rejection entails the identification of outliers and then their elimination from the data set. Robust procedures are the primary accommodation approach and allow statistical inferences to be made that are tolerant of outliers.

Given that DEA methods depend upon 'extreme' points to define a frontier, simply eliminating outliers from the data set may not be a useful approach. The application of robust procedures has been applied to DEA. Seaver and Triantis (1995) and Seaver and Triantis (1989) investigate the impact of outliers and leverage points on technical efficiency measurement using robust procedures. In order to detect outliers and leverage points the

data is modeled as a standard linear relationship. Robust procedures are then applied. This technique is applied to the evaluation of the efficiency of three paper mills using monthly production data over three years. Finding leverage points and outliers can help identify the observations during which the mills experience unusual production situations. Such unusual observations are of particular interest is formulating efficiency performance improvement strategies. The result of the analysis is that outlier and leverage points should not be eliminated from the dataset unless they represent serious measurement errors. The analysis also implies that the "accuracy of the computed efficiency measures hinges in part on the varying production information these observations [outliers and leverage points] represent for each mill" (Seaver and Triantis, 1991, p. 26).

O.B. Olesen and N.C. Peterson (1991, 1993) present a method for constructing a DEA frontier with only Full Dimensional Facets (meaning facets are composed of points defined by all inputs and outputs). Non-Full Dimensional Facets essentially restrict the substitution possibilities among inputs and outputs and so are eliminated. This method is suggested to be appropriate if a "fixed" frontier technology is unacceptable. Testing for Non-Full Dimensional Facets also has the advantage of flagging a lack of data variation; which DEA methods do not. Where insufficient data variation exists aggregation of inputs and outputs may be done to obtain sufficient data variation. Olesen and Peterson (1993) apply the condition that for an aggregation to be admissible that no efficient DMU, in the original space, is termed inefficient after aggregation.

Wilson (1995) presents a technique for detecting influential observations in Data Envelopment Analysis. The purpose is to identify observations that should be more closely scrutinized. This is done by metrics based on the elimination of an ostensibly efficient observation from the set of observations and then re-calculating efficiency measures. If there is a significant change then the eliminated observation is deemed influential.

Since approaches to the measurement of efficiency in some fashion partition a set of production plans, methods of data analysis from cluster analysis have applicability. The

"clusters" are already defined by methods of efficiency analysis. Thanassoulis (1996) clusters DMU's according to the output mix. Similar DMU's making-up a cluster can then be evaluated in terms of resource allocation matching operating priorities and exogenous factors. Seaver and Triantis (1992) apply a two stage procedure to define clusters - first the Kth nearest neighbor approach to define clusters and then define fuzzy sets based on the initial cluster sets. Fuzzy sets are applied because clustering techniques produce hard placement of observations in a cluster. Some observations in efficiency analysis will not necessarily fall into the extremes of efficient and inefficient behavior but somewhere in between. This approach was applied to data from three pulp and linerboard facilities and it was found that this fuzzy K - means clustering technique allows the analyst to obtain some understanding of the degree of belonging to derived efficiency clusters and that observations that represent extreme of unusual operating conditions were for the most part highlighted (Seaver and Triantis, 1992).

Cluster analysis as used in standard statistical analysis is also applicable to the research. Cluster analysis distance measures are reviewed based on Lorr (1983). A cluster of points may be characterized by three components: elevation (level), scatter, and shape. "The elevation component Xi is the mean of all scores for an entity. Scatter Si is defined as the square root of the sum of squares of the entity's deviation scores around its own mean. Shape is the information remaining in the score after removing elevation and equilizing scatter" (Lorr, 1983, p. 28). Raw Scores may be converted to Deviation Scores by subtracting out the mean of the Raw Score for each measurement. Standard Scores are obtained by reducing all vectors to unit length from the origin. The meaning of Shape for the measurement of productive efficiency is not clear. However, elevation and scatter measurements can provide useful information. Distance measures presented by Lorr (1983) are:

$$D_{ih}^2 = \sum_{j=1}^{k} (X_{ij} - X_{hj})^2 \qquad \text{(Eq. 2-45)}$$

Where,

i and h are two points.

j = 1,…,k the variables.

Using Deviation Scores another distance measure between two points i and h is:

$$D_{ih}^2 = \sum x_{ij}^2 + \sum x_{hj}^2 - 2\sum x_{ij}x_{hj} + k(\overline{X}_i - \overline{X}_h)^2 \qquad \text{(Eq. 2-46)}$$

where,

$\overline{X}$ = the mean for the distance measure or scores of a point.

Another metric could simply be the sum of the absolute value of the difference.

$$D_{ih} = \sum_{j=1}^{k} \left| X_{ij} - X_{hj} \right| \qquad \text{(Eq. 2-47)}$$

Aldenderfer and Blashfield (1984, p. 26) summarize the problems with distance measures in cluster analysis:

> "[D]istance measures suffer from serious problems, among the most critical of which is that the estimation of the similarity between cases is strongly affected by elevation differences. Variables with both large size differences and standard deviations can essentially swamp the effects of other variables with smaller absolute sizes and standard deviations. Moreover, distance metrics are also affected by transformations of the scale of measurement variables…."

Kaufman and Rousseeuw (1990, p. 5) note that "changing the measurement units may even lead one to see a very different clustering structure." It is also noted that the units of measurement are essentially assigning relative weights to variables. One way to avoid the problem of unit dependence in the measurement of distance is to standardize input. A method presented by Kaufman and Rousseeuw (1990, pp. 8-9) is to first calculate the average for each variable, calculate a metric of dispersion, and then the standardized score. These formulas for n objects (or observations) and p variables are:

$$m_f = \frac{1}{n}(x_{1f} + x_{2f} + ... + x_{nf}) \qquad \text{(Eq. 2-48)}$$

for each f=1,…,p

The dispersion is calculated as:

$$s_f = \frac{1}{n}\{|x_{1f} - m_f| + |x_{2f} - m_f| + ... + |x_{nf} - m_f|\} \qquad \text{(Eq. 2-49)}$$

This measure of dispersion is preferred to the standard deviation since it is more robust in terms of not changing that much if there is one outlying observation.

$$z_{if} = \frac{x_{if} - m_f}{s_f} \qquad \text{(Eq. 2-50)}$$

where i = 1,…n. All of the variables are replaced with values of z_{if} which is a unitless measure with a mean of zero and an absolute deviation of 1.0.

2.3.6 *Aggregation*

Charnes, Cooper, and Rhodes (1978) suggest that the DEA method, since it derives from empirical observation, bypasses the intractable problem of aggregations associated with production functions. Shephard (1981) concludes that for the Cobb-Douglas production function the aggregate marginal productivity's may be interpreted as averages of the corresponding derivatives of the micro-economic production function. More recent research indicates that aggregation does affect the measurement of productive efficiency.

The issue of aggregation is important in economics since many economic indicators are based on the aggregation of data. One bias that results from the aggregation of data for an economic indicator is the time lag involved in compiling the data so that the resulting indicator is based on data that does not necessarily represent the true state of the system (Oguchi, 1990) (Goodfriend, 1992). Other issues in economics are whether to use micro or macro equations to predict aggregate variables and the determination of aggregation bias "defined by the deviation of the macro parameters from the average of the corresponding micro parameters" (Lee, 1990). Primont (1983) addresses the aggregation of input price subindexes and shows that aggregation into a complex index requires that groups of prices for subindexes must be separable in the cost function. Primont (1993) again addresses this issue as it relates to efficiency measures. A set of inputs is evaluated for overall, technical, and allocative efficiency for an unaggregated set and an aggregated set (obtained by partitioning and then aggregating inputs in the subsets). Primont (1993) finds that overall efficiency measures are unbiased with input-cost aggregates and all inputs variable.

Technical efficiency is biased downward with allocative efficiency being biased upwards. Where some inputs are fixed overall efficiency (as well as technical and allocative efficiency) are biased. Primont (1993) suggests that the most disaggregated data available should be used where possible to avoid bias in efficiency measures.

The impact of input aggregation on technical efficiency is investigated by Thomas and Tauer (1994). Typically, it is assumed that the aggregating function is the linear aggregator weighted by prices. Using data from New York dairy farms, Thomas and Tauer (1994), show that relative efficiency changes dramatically as inputs are aggregated. Specifically, it is shown that linear input aggregation can reduce the measured technical efficiency of any given firm.

2.3.7   *Ordinal Inputs and Outputs*

Golany (1988) incorporates ordinal relations among inputs by sequentially accumulating the relevant inputs - i.e., adding input over time. Iqbal, Wade, and Seiford (1991) further develop the approach to include a strong ordinal relationship (strict linear ranking by weights) and weak ordinal relationship (ties are permissible). Cook and Kress (1990) apply DEA to determine weights used to aggregate voter preference rankings. A candidate has so many first place votes, second place votes, etc. A DEA formulation is used to determine weights assigned to the first place votes, second place votes, and so on. These weights are then used to aggregate the scores into a single score. Zhu (1996) modifies DEA analysis by introducing weights for inputs and outputs (in addition to the weights assigned by the DEA procedure). The constraints stay the same, but in one formulation presented for a non-radial CCR model the objective function for the input reducing model, $\min \frac{1}{I} \sum_{i=1}^{I} h_i$, is

modified to, $\min \dfrac{\sum_{i=1}^{I} A_i h_i}{\sum_{i=1}^{I} A_i}$. As noted by Zhu, there are practical difficulties in assigning the

weights, A_i.

Banker and Morey (1986) incorporate categorical variables into DEA. This is accomplished by adding an indicator variables (that takes on value of 0 or 1) to the standard DEA formulation. This allows inputs where the value is say low, medium, or high to be incorporated into the DEA formulation. This approach is more applicable to the comparison of different DMU's where there may be some qualitative differences (e.g., one bank has a drive through and another does not). Olesen and Peterson (1995) incorporate quality of outputs as measured on an ordinal scale into DEA analysis. A value judgment, in terms of quality, is essentially imposed on outputs. Three approaches are used as described below (Oleson and Peterson, 1995, p. 118).

1) Each output is disaggregated into different types of output according to the cumulative quantities of the output produced at or above a given level of quality.
2) The cumulative probabilities for each output to be produced at or above any given level of quality are included as a characteristic of the production along with quantities produced. Quality is said to increase if the cumulative probability for a given output to be produced at or above any given level of quality increases for at least one level of quality and does not decrease for any level of quality.
3) The cumulative probabilities for each output to be produced at of above any given level of quality can be seen as the only characteristics of the production, i.e. quantities produced are ignored. A DMU is termed inefficient if there exits a convex combination of DMUs with a cumulative probability distribution at or above any level of quality for all outputs.

Cooper, Park, and Yu (1998) take previous work for incorporating imprecise data into DEA and provide an overall framework for formulating DEA linear programs that have bounds and ordinal classifications.

2.4   Application of Production Theory to Pollution Prevention and Control

Many of the various methods of measuring productive efficiency have been applied to the consideration of pollution control. In the measurement of productive efficiency waste products are sometimes explicitly considered and sometimes not. Both econometric and DEA based techniques have been used to evaluate the affect of pollution controls on

productive performance. There are three approaches, with many variations, that have been used. One is to apply standard DEA or econometric analysis to a production system over time or to different sets of production systems (with and without pollution controls) and then compare the results from which estimates of the cost in terms of lost production associated with pollution controls can be estimated. The other two approaches explicitly consider pollution or undesirable outputs and determine the cost of pollution control based on loss of disposability or assign shadow prices to the undesirable outputs. These three approaches are each reviewed.

2.4.1 *Econometric Approaches*

Ethridge (1972) modifies the standard economic approach to cost and net revenue functions to include waste. Outputs are classified as primary product, Y_p, and a side product, Y_0. All Y_0 is either recovered as by-products, Y_b, or are discarded, Y_w. Traditional analysis has the firm producing Y_p and Y_b according to cost and prices with Y_w imposing an external cost on society. A constant k is assumed such that $Y_w = k Y_p - Y_b$. As noted by Ethridge (1972) a change in k can represent new technology, changes in input mix, and production methods. With the further assumption of a charge for waste disposal capacity of the resource the costs of waste may be included in total cost and net revenue functions. The implication of the analysis is that in response to increased user charges on waste disposal a firm responds by increasing by-product recovery, Y_b, and decreasing primary product, Y_p, and /or decreasing k.

Tran (1982) applies an econometric approach that explicitly considers emissions controlled by regulation, (i.e., particulates, sulfur dioxide, and oxides of nitrogen, change in cooling water temperature), along with electricity in the formulation of a cost function. Results of the analysis of a data set of 173 plants indicate that the cost functions are not multiplicatively separable in outputs; in other words, a joint consideration of waste outputs is justified in the cost function if elasticities of factor substitution are to be accurately estimated. This indicates that the explicit consideration of undesirable outputs is necessary

for an accurate representation of a technology.   Gallop (1983) developed a similar econometric cost model to show that productivity in fossil-fueled electric power generating plants has declined as a result of regulations on sulfur dioxide emissions. Jastrow (1985) develops an econometric model for the construction cost that explicitly considers the cost of pollution control equipment such as scrubbers and also has a general productivity term that includes costs associated with increasingly stringent environmental controls.   Results indicate significant productivity decline due to environmental controls, but these results are ambiguous since productivity declines begin before the imposition of many environmental regulations.   Clearly there are other factors at work in declining productivity of plant construction that are not identified in the model.

In a linear programming approach, Berstein (1990), selects the most efficient plants from a sample of coal-fired power plants to develop a regression model specifying the frontier. Waste is not explicitly considered in the model. Instead, the effect of pollution controls is evaluated by measuring the efficiency of plants with and without scrubbers. Results indicate a decline in efficiency where scrubbers are employed.

2.4.2   *Data Envelopment Analysis Approaches*

Tyteca (1995) compares four methods of evaluating environmental performance to data from power plants. The methods are those developed by Färe, Grosskopf, Knox Lovell, and Pasurka (1989) (output oriented, and both input and output oriented); Hayes, Ratick, and Cummings-Saxton (1993); and a fourth approach based simply on the ratios of undesirable outputs to desirable outputs. Significant variations were found in the relative efficiencies of the power plants evaluated.   Tyteca suggests that this provides complementary information.   Tyteca (1996) reviews methods for measuring the environmental performance of firms with an emphasis on DEA methods. Tyteca suggest that a best practice frontier could be defined (even if this frontier were not practically achievable) as a measure of relative environmental performance. For this reason, FDH is

rejected as an approach for measuring environmental performance since the use of a more "abstract" frontier is desirable.

The two main variations on the standard DEA approach for the incorporation of environmental considerations are disposability and shadow prices. These approaches are reviewed in the next two sections.

2.4.2.1 *Disposability*

The assumption of strong disposability means that an undesirable output can be disposed of without cost. The assumption of weak disposability means that at least some input must be used to dispose of an undesirable output, thereby, reducing the desirable outputs. For the estimation of the loss of disposability, Färe (1983) defines three reference technology subsets for measuring output efficiency based on variations in the assumption of disposability. These are[1]:

$$\text{Isoq } P(x) := \quad \{u: u \in P(u), \Theta\, \mathbf{u} \notin P(x) \text{ for } \Theta > 1 \} \qquad \text{(Eq. 2-51)}$$
$$P(x) \neq \{0\}.$$
$$\{0\}, P(x) = \{0\}.$$

$$\text{Weff } P(x) := \{u: u \in P(x), v >^* \ u \Rightarrow v \notin P(x)\} \qquad \text{(Eq. 2-52)}$$
$$P(x) \neq \{0\}.$$
$$\{0\}, P(x) = \{0\}.$$

$$\text{Eff } P(x) := \quad \{u: u \in P(x), v \geq u \Rightarrow v \notin P(x)\} \ (u > \mathbf{o}) \qquad \text{(Eq. 2-53)}$$
$$P(x) \neq \{0\}.$$
$$\{0\}, P(x) = \{0\}.$$

These subsets are the output isoquant, output weak efficient, and output efficient. Färe then defines a measure of congestion, or lost output due to the lack of disposability to be: C(x,u)

---

[1] The term x >* y means if and only if $x_n > y_n$ or $x_n = y_n$, n = 1, 2, …,N.

= W(x,u)/F(x,u). The Farrell Output Efficiency Measure F(x,u) associated with Isoq P(x) is defined to be:

$$F(x,u) := \max \{ \Theta \geq 0 : \Theta \cdot u \in P(x)\}, \qquad (Eq.\ 2\text{-}54)$$
$$(x,u) \in D(F),$$
$$+\infty, \quad (x,u) \in Complement\ D(F).$$
where,
$$D(F) := \{(x,u): \exists\ \Theta \geq 0\ such\ that\ \Theta \cdot u \in P(x)\}$$

The Weak Output Efficiency Measures associated with Weff P(x) is defined to be:

$$W(x,u) := \max \{ \Theta \geq 0 : \Theta \cdot M(u) \cap P(x) \neq \varnothing \}, \qquad (Eq.\ 2\text{-}55)$$
$$(x,u) \in D(W),$$
$$+\infty, \quad (x,u) \in Complement\ D(W).$$
Where,
$$M(u) := \{v: v \geq u\},$$
$$D(W) := \{(x,u): \exists\ \Theta \geq 0\ such\ that\ \Theta \cdot M(u) \cap P(x) \neq \varnothing \},$$

F(x,u) will always be less than or equal to W(x,u). W(x,u) represents the measure of efficiency relative to the unregulated technology production frontier. F(x,u) represents the measure of efficiency relative to the regulated technology production frontier. The measure of efficiency loss due to environmental control regulations is C(x,u). "By comparing these efficiencies for both weak and strong disposability (frontier) technologies, the indirect impact of environmental control on relative output efficiency is calculated. Färe (1986) shows that the DEA method of CCR determines an input vector efficient if and only if it belongs to the weak efficient subset.

A variation on this free disposability/weak disposability ratio approach was applied by Brännlund, Färe, and Grosskopf (1995) where a short-run quantity constrained (by environmental permit restriction on undesirable outputs) profit function and a short-run unregulated profit function are calculated for Swedish Pulp and Paper Mills Plant level data in 1989 and 1990. A measure of the cost of regulation is determined by the ratio of the profit functions.

Färe and Grosskopf (1995) build on the work of Hayes, Ratick, Bowen and Cummings-Saxton (1993) to include waste outputs explicitly in a DEA measure for decision

making associated with environmental protection. It is shown that "the jointness between desirable and undesirable outputs can be incorporated into environmental decision models" (Färe and Grosskopf, 1995, p. 1). The concept of nulljointness is used where the subvector y is nulljoint with z if for all $(y,z) \in P(x)$ and $x \in \mathfrak{R}_+^N$, z=0 implies y=0. If there are no undesirable outputs then there can be no desirable outputs.

Outputs in the Färe and Grosskopf model (1995) are divided into sets of 'good,' and 'bads' and defined to be:

$$P(x) = \{(y,z): \sum_{k=1}^{K} \lambda_k y_{km} \geq y_m, \ m=1,\dots M, \qquad\qquad \text{(Eq. 2-56)}$$

$$\sum_{k=1}^{K} \lambda_k z_{k\ell} = z_\ell, \ \ell = 1,\dots, L,$$

$$\sum_{k=1}^{K} \lambda_k x_{ki} \leq x_n, \ i=1,\dots, I,$$

$$\lambda_k \geq 0, \ k=1,\dots, K\}$$

where,

$y = (y_1, \dots, y_m) \in \mathfrak{R}_+^M$ are good outputs.

$z = (z_1, \dots, z_L) \in \mathfrak{R}_+^L$ are bad outputs

$x = (x_1, \dots, x_I) \in \mathfrak{R}_+^I$ are the inputs

$k = 1, \dots K$ observations

Note that this notation has diverged from the standard of specifying an output by u as a convenience since there are now two classes of outputs. The constraint for the 'bad' outputs, z, differs from the standard DEA formulation in that these outputs are restricted from increasing. The set, P(x), satisfies Constant Return to Scale and Weak Disposability of outputs. Färe and Grosskopf (1995) suggest a measure of efficiency "that maximizes each desirable output and at the same time minimizes each undesirable output and also accounts for all inputs" (Färe and Grosskopf, 1995, p. 11). This measure is:

$$E(x^{k'}, z^{k'}, y^{k'}) = \max \ \sum_{k=1}^{\hat{K}} \frac{\theta_m}{\hat{K}} \ - \ \sum_{k=1}^{\tilde{K}} \frac{\mu_\ell}{\tilde{K}} \qquad\qquad \text{(Eq. 2-57)}$$

$$\text{s.t.} \qquad \theta_m y_{k'm} \ \leq \ \sum_{k=1}^{K} \lambda_k y_{km}, \ \ m=1, \ , M,$$

$$\mu_\ell z_{k'\ell} \ = \ \sum_{k=1}^{K} \lambda_k z_{k\ell} , \ \ \ell=1, ..., L,$$

$$\sum_{k=1}^{K} \lambda_k x_{ki} \ \leq \ x_{k'i}, \ i=1,..., I,$$

$$\lambda_k \ \geq \ 0, \ k=1, ..., K.$$

where,

$\hat{K}$ is the number of positive $y_{k'm}$

$\tilde{K}$ is the number of positive $z_{k'\ell}$

Two terms are required in the objective function since the solution procedure seeks to both maximize desirable outputs and minimize undesirable outputs. This measure allows the bad and good outputs to be increased at different rates - i.e., the measure is nonradial.

2.4.2.2 *Shadow Prices*

The linear programming formulation of data envelopment analysis lends itself to the use of shadow prices for estimating the cost of undesirable outputs. The standard or primal formulation of a linear program is transformed into the dual according to the following rules (Taha, 1982, p. 106):

1. For every primal constraint there is a dual variable.
2. For every primal variable there is a dual constraint.
3. The constraint coefficients of a primal variable form the left-side coefficients of the corresponding dual constraints; and the objective coefficients of the same variable becomes the right side of the dual constraint.

The optimal solution of the primal can be obtained from the dual and visa versa. The economic interpretation of shadow prices is not straightforward.

Estimating shadow prices for undesirable outputs is the other primary approach used in DEA analysis to account for loss of efficiency from regulatory restrictions. Pittman (1983) estimates shadow prices as part of a modification to the index proposed by Caves, Christensen and Diewert (1992). A data point of k outputs is decomposed into desirable output, Y^k_1 and undesirable outputs, Y^k_2, and Y^k_3 (say water and air pollution). The factor of proportionality δ_k is redefined to be "the minimum proportional decrease in Y^k_1 and increase in Y^k_2, and Y^k_3 such that this combination of output is producible with the input and productivity levels of [another data point]" (Pittman, 1983, p. 885). Since the Caves, Christensen, and Diewert index assumes constrained profit maximization, some revenue share must be assigned to the undesirable outputs. Pittman (1983) assigns shadow prices for a data set of paper mills based on calculations from two data sets that assign costs to the level of pollution abatement.

Atkinson and Halvorsen (1984) propose a generalized cost function that does not require the assumption of cost minimization (or profit maximization). Instead, it is assumed that production decisions are based on "unobservable shadow prices which reflect the effects of regulation on the effective prices of inputs" (Atkinson and Halvorson, 1984, p. 647).

Färe (1990) introduces an approach to estimating shadow prices that "identifies separate vectors of relative shadow prices for each observation" (Färe, 1990, p. 710). The approach is based on the assumption that at least one input market is efficient. Where L(u) is convex and inputs are disposable, there exists a shadow price vector, p*, such that p*x minimizes cost for a given output level u.

2.5   Relevance of the Reviewed Literature

A wide variety of literature was reviewed in Chapter 2 as the starting point for this research. There are four main areas of literature that were reviewed. First, the context and motivation for the research is provided in terms of other efforts to integrate the often conflicting objectives of environmental protection and economic performance through increased productive efficiency. One common tread in these reviewed methods is the treatment of a production process as a "black box." In other words, they are techniques that are largely applied by economists to study entire industries or to perform highly complex studies that attempt to quantify environmental impacts. While these approaches are certainly useful in some contexts, they do not seem well suited to performance measurement at the operational level of a single manufacturing facility; which is the purpose of this research.

Second, an area that has been applied to some extent to operational data is the measurement of productive efficiency. Standard methods of measuring productive efficiency were reviewed in Section 2.2. The concepts presented as part of this review are the basis for the environmental performance metrics proposed in Chapter 3. While these standard methods differ widely in mathematics, they all share some basic elements. Some reference against which productive efficiency is measured must be selected. Then, some method of measuring the distance of a particular production plan to the reference is implemented to obtain a measure of efficiency. This basic approach is also used in the research. Of the methods reviewed in Section 2.2 the Benchmark Correspondence method is the closest to the approach that is developed as part of this research. Benchmark Correspondence differs from the other standard methods in that the reference is actual data; not a function or frontier defined based on data. This means that the distance measure is from one production plan to another and not from a production plan to a reference function or frontier that is defined mathematically.

Third, a wide variety of approaches and methods are then reviewed that relate to the research. Multi-objective programming is reviewed since this technique deals with attempting to reconcile sometimes conflicting objectives. Input substitution, particularly where there was consideration of environmental impacts, is reviewed since substitution is one method of reducing environmental impacts. Attempts that have been made to measure environmental performance at the manufacturing facility level are then reviewed. These approaches seem to be generally lacking in theoretical foundation and often require large amounts of data that is not available. Since data was obtained from a manufacturing facility for this research a review of cost accounting was performed to obtain insight into data requirements. The topics of data analysis and aggregation are also reviewed in Section 2.3 since these are relevant to the treatment of the data. Finally, the issue of ordinal ranking is reviewed since the detail of data available to a manufacturing facility will likely only allow such ranking; rather than detailed estimates of relative environmental benefits or impacts.

Fourth, Section 2.4 presents research that has attempted to modify standard methods of measuring productive efficiency with the explicit inclusion of undesirable outputs. These methods share the general applicability of standard methods of measuring productive efficiency to industry level studies. The value and meaning of these measures at the manufacturing facility operational level are not clear.

The method presented in Chapter 3 brings together these concepts in an attempt to provide a method of measuring environmental performance that is theoretically founded and also appropriate for the types of data typically available at a manufacturing facility.